FORM PTO-1390 (REV. 12-2001)	U.S. DEPARTMENT OF COMMERCE PATENT AND TRADEMARK OFFICE	ATTORNEY'S DOCKET NUMBER S-31147A

TRANSMITTAL LETTER TO THE UNITED STATES DESIGNATED/ELECTED OFFICE (DO/EO/US) CONCERNING A FILING UNDER 35 U.S.C. 371

U.S. APPLICATION NO. (If known, see 37 CFR 1.5

10/088384

INTERNATIONAL APPLICATION NO. PCT/EP00/08994	INTERNATIONAL FILING DATE September 14, 2000	PRIORITY DATE CLAIMED September 16, 1999

TITLE OF INVENTION
Transcriptionally Silenced Plant Genes

APPLICANT(S) FOR DO/EO/US
Steimer, Andrea et al.

Applicant herewith submits to the United States Designated/Elected Office (DO/EO/US) the following items and other information:

1. [X] This is a **FIRST** submission of items concerning a filing under 35 U.S.C. 371.

2. [ ] This is a **SECOND** or **SUBSEQUENT** submission of items concerning a filing under 35 U.S.C. 371.

3. [x] This is an express request to begin national examination procedures (35 U.S.C. 371(f)). The submission must include items (5), (6), (9) and (21) indicated below.

4. [ ] The US has been elected by the expiration of 19 months from the priority date (Article 31).

5. [x] A copy of the International Application as filed (35 U.S.C. 371(c)(2))
     a. [x] is attached hereto (required only if not communicated by the International Bureau).
     b. [ ] has been communicated by the International Bureau.
     c. [ ] is not required, as the application was filed in the United States Receiving Office (RO/US).

6. [ ] An English language translation of the International Application as filed (35 U.S.C. 371(c)(2)).
     a. [ ] is attached hereto.
     b. [ ] has been previously submitted under 35 U.S.C. 154(d)(4).

7. [ ] Amendments to the claims of the International Aplication under PCT Article 19 (35 U.S.C. 371(c)(3))
     a. [ ] are attached hereto (required only if not communicated by the International Bureau).
     b. [ ] have been communicated by the International Bureau.
     c. [ ] have not been made; however, the time limit for making such amendments has NOT expired.
     d. [ ] have not been made and will not be made.

8. [ ] An English language translation of the amendments to the claims under PCT Article 19 (35 U.S.C. 371 (c)(3)).

9. [x] An oath or declaration of the inventor(s) (35 U.S.C. 371(c)(4)).

10. [ ] An English lanuage translation of the annexes of the International Preliminary Examination Report under PCT Article 36 (35 U.S.C. 371(c)(5)).

Items 11 to 20 below concern document(s) or information included:

11. [ ] An Information Disclosure Statement under 37 CFR 1.97 and 1.98.

12. [x] An assignment document for recording. A separate cover sheet in compliance with 37 CFR 3.28 and 3.31 is included.

13. [x] A FIRST preliminary amendment.

14. [ ] A SECOND or SUBSEQUENT preliminary amendment.

15. [ ] A substitute specification.

16. [ ] A change of power of attorney and/or address letter.

17. [x] A computer-readable form of the sequence listing in accordance with PCT Rule 13ter.2 and 35 U.S.C. 1.821 - 1.825.
     + Paper Copy of Sequence Listing.

18. [ ] A second copy of the published international application under 35 U.S.C. 154(d)(4).

19. [ ] A second copy of the English language translation of the international application under 35 U.S.C. 154(d)(4).

20. [X] Other items or information: **Certified copy of Priority Document GB9921964.4**

Application Data Sheet; International Search Report; PTO-2038

U.S. APPLICATION NO. (if known, see 37 CFR 1.5) 10/088,584	INTERNATIONAL APPLICATION NO. PCT/EP00/08994	ATTORNEY'S DOCKET NUMBER S-31147A

	CALCULATIONS PTO USE ONLY
21. [x] The following fees are submitted:	

BASIC NATIONAL FEE (37 CFR 1.492 (a) (1) - (5)):

Neither international preliminary examination fee (37 CFR 1.482) nor international search fee (37 CFR 1.445(a)(2)) paid to USPTO and International Search Report not prepared by the EPO or JPO $1040.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but International Search Report prepared by the EPO or JPO $890.00

International preliminary examination fee (37 CFR 1.482) not paid to USPTO but international search fee (37 CFR 1.445(a)(2)) paid to USPTO $740.00

International preliminary examination fee (37 CFR 1.482) paid to USPTO but all claims did not satisfy provisions of PCT Article 33(1)-(4) $710.00

International preliminary examination fee (37 CFR 1.482) paid to USPTO and all claims satisfied provisions of PCT Article 33(1)-(4) $100.00

ENTER APPROPRIATE BASIC FEE AMOUNT =	$ 890.00	

Surcharge of **$130.00** for furnishing the oath or declaration later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(e)).	$	

CLAIMS	NUMBER FILED	NUMBER EXTRA	RATE	$	
Total claims	16 - 20 =		x $18.00	$	
Independent claims	2 - 3 =		x $84.00	$	
MULTIPLE DEPENDENT CLAIM(S) (if applicable)			+ $280.00	$	
TOTAL OF ABOVE CALCULATIONS =				$	
[ ] Applicant claims small entity status. See 37 CFR 1.27. The fees indicated above are reduced by 1/2. +				$	
SUBTOTAL =				$	
Processing fee of **$130.00** for furnishing the English translation later than [ ] 20 [ ] 30 months from the earliest claimed priority date (37 CFR 1.492(f)).				$	
TOTAL NATIONAL FEE =				$	
Fee for recording the enclosed assignment (37 CFR 1.21(h)). The assignment must be accompanied by an appropriate cover sheet (37 CFR 3.28, 3.31). **$40.00** per property +				$ 40.00	
TOTAL FEES ENCLOSED =				$ 930.00	

	Amount to be refunded:	$
	charged:	$

a. [ ] A check in the amount of $ _____ to cover the above fees is enclosed.

b. [ ] Please charge my Deposit Account No. _____ in the amount of $ _____ to cover the above fees. A duplicate copy of this sheet is enclosed.

c. [X] The Commissioner is hereby authorized to charge any additional fees which may be required, or credit any overpayment to Deposit Account No. 50-1744 . A duplicate copy of this sheet is enclosed.

d. [X] Fees are to be charged to a credit card. **WARNING: Information on this form may become public. Credit card information should not be included on this form.** Provide credit card information and authorization on PTO-2038.

NOTE: Where an appropriate time limit under 37 CFR 1.494 or 1.495 has not been met, a petition to revive (37 CFR 1.137 (a) or (b)) must be filed and granted to restore the application to pending status.

SEND ALL CORRESPONDENCE TO:

Customer Number 22847

Marcia R. Morton
SIGNATURE

Marcia R. Morton
NAME

46,942
REGISTRATION NUMBER

S-31147A

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	
STEIMER *et al.*	Art Unit: TBA
Serial Number: National Stage Application of PCT/EP00/08994	Examiner: TBA
Filed: March 15, 2002	Atty Docket: S-31147A
For: Transcriptionally Silenced Plant Genes	

PRELIMINARY AMENDMENT

BOX PCT
U.S. Patent and Trademark Office
P.O. Box 2327
Arlington, VA 2202

Sir:

Applicants respectfully request that the above-captioned application be amended as follows in advance of prosecution:

IN THE SPECIFICATION:

At page 1, after the title, please insert:

This application is a § 371 of International Application No. PCT /EP00/08994, filed September 14, 2000.

IN THE CLAIMS

Please amend claim 5 to read as follows:

5.   (Amended) The nucleic acid of claim 1 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

Please add the following new claims:

14. (New) The nucleic acid of claim 2 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids length in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

15. (New) The nucleic acid of claim 3 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

16. (New) The nucleic acid of claim 4 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

REMARKS

Claim 5 has been amended and new claims 14-16 have been added to eliminate multiple dependency and to ensure that they are in compliance with the rules of U.S. practice. No new matter has been added.

If in the opinion of the Examiner, a telephone conference would expedite the prosecution of the subject application, the Examiner is invited to call the undersigned attorney.

The Commissioner has been authorized to charge the fees associated with the filing of the application that this amendment accompanies and associated papers to Applicant's credit card. If there are any additional fees required, authorization is hereby provided to

charge such fees to Applicant's Deposit Account No. 50-1744 (in the name of Syngenta Biotechnology, Inc.).

Respectfully submitted,

Syngenta Biotechnology, Inc.   _Marcia R. Morton_
Patent Department   Marcia R. Morton
P.O. Box 12257   Attorney for Applicants
Research Triangle Park, NC 27709-2257   Reg. No. 46,942
(919) 541-8566
December 21, 2001

Version With Markings To Show Changes Made

The specification has been amended as follows to include the priority data:

--This application is a § 371 of International Application No. PCT /EP00/08994, filed September 14, 2000.--

The claims have been amended as follows:

5. (Amended) The nucleic acid of [any one of claims 1 to 4] claim 1 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

-- 14. (New) The nucleic acid of claim 2 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids length in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

15. (New) The nucleic acid of claim 3 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.

16. (New) The nucleic acid of claim 4 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids in length being at least 85% identical to an aligned component of SEQ ID NO: 10.--

Transcriptionally Silenced Plant Genes

The present invention relates to the field of gene expression in plants and in particular concerns gene silencing, a phenomenon frequently observed after integration of transgenes into plant genomes. Comparison of transcriptional gene expression between an Arabidopsis line carrying a silent transgene present in multiple copies and its mutant derivative *mom1* impaired in silencing of the transgene revealed two cDNA clones which are expressed in the mutant plants, but not in the parental and not in wild type plants. Both clones are derived from the same family of transcripts which we refer to as TSI (Transcriptionally Silent Information). The disclosed genomic templates encoding TSI are repetitive elements with mainly pericentromeric location and conserved organization among various ecotypes. They are also referred to as TSI. Transcriptional silencing of the genomic TSI templates is specifically released in the mutant. Silencing of said templates is further released in other genotypes known to affect transcriptional gene silencing. Thus, transcription of TSI can be used as a marker to identify a defective silencing pathway in a plant.

Correct balance between activation and silencing of its genetic information is essential for any living cell. A tight control of gene expression is necessary for adaptation to environmental factors, regulation of physiological requirements, and development of differentiated, specialized cell types within a multi-cellular organism. For example differentiation processes involve mitotically heritable changes of gene expression, wherein the acquired states of gene activity gain a certain stability. This stability can be achieved by the strict control of gene activators, by regulation of transcript stability, or by regulating the transcriptional availability of genetic information itself as by stable silencing of selected genetic loci. Silencing has been frequently observed in connection with repression of transgene expression in various experimental systems.

In plants, silencing of transgenic loci limits the reliability of transgenic approaches to improve quality traits. It has been noticed that complex inserts containing rearranged multiple copies of a transgene are particularly prone for gene silencing. Two different mechanisms leading to loss of transgene expression are observed. The first prevents transcription (transcriptional gene silencing or TGS), and the second targets selected transcripts for rapid degradation (posttranscriptional gene silencing or PTGS). Triggers of both processes seem to be similar, since the onset of both types of silencing correlates with redundancy of genetic information, i.e. DNA repeats in case of TGS and RNA

overproduction for PTGS. TGS is meiotically heritable and correlates with DNA template modification manifested by hypermethylation of promoters of silenced genes or with local changes of chromatin structure. In contrast, PTGS is not meiotically transmitted and needs to be reestablished in each sexual generation. PTGS does not require modification of a DNA template, however, increased levels of DNA methylation within the protein-coding region of silenced genes have been observed.

The majority of silencing studies in plant systems deal with silencing of transgenes. There are only a few examples of gene silencing without involvement of transgenic loci. The criteria for TGS susceptibility of genetic information is very poorly understood, and the natural targets of transcriptional silencing in a normal, wild type plant are yet to be discovered. It has been postulated that TGS is a defense system against invasive DNA such as transposable elements but experimental evidence for this hypothesis is lacking.

Within the context of the present invention reference to a gene is to be understood as reference to a DNA coding sequence associated with regulatory sequences, which allow transcription of the coding sequence into RNA such as mRNA, rRNA, tRNA, snRNA, sense RNA or antisense RNA. Examples of regulatory sequences are promoter sequences, 5' and 3' untranslated sequences, introns, and termination sequences.

A promoter is understood to be a DNA sequence initiating transcription of an associated DNA sequence, and may also include elements that act as regulators of gene expression such as activators, enhancers, or repressors.

Expression of a gene refers to its transcription into RNA or its transcription and subsequent translation into protein within a living cell.

Any part or piece of a specific nucleotide or amino acid sequence is referred to as a component sequence.

It is the aim of the present invention to provide nucleic acid molecules encoding genetic information which is not expressed, i.e. silenced, in wild type plants but whose expression is turned on in plants which are defective in transcriptional gene silencing. Said molecules can be defined by the formula R_A-R_B-R_C, wherein

-- R_A, R_B and R_C indicate component sequences consisting of nucleotide residues
      independently selected from the group of G, A, T and C or G, A, U and C, wherein
      G is Guanosinmonophosphate,
      A is Adenosinemonophosphate,

T is Thymidinmonophosphate,

U is Uridinmonophosphate and

C is Cytidinmonophosphate;

-- R_A and R_C consist independently of 0 to 6000 nucleotide residues;

-- R_B consists of at least 50 nucleotide residues; and

— the component sequence R_B is at least 80% identical to an aligned component
sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID
NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

In a preferred embodiment of the present invention R_B consists of at least 100 nucleotide
residues and is at least 85% identical to an aligned component sequence of SEQ ID NO: 1,
SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7,
SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

In another preferred embodiment R_B consists of at least 200 nucleotide residues and is at
least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2,
SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8,
SEQ ID NO: 9 or SEQ ID NO: 27.

Specific examples of R_B are the sequences given in SEQ ID NO: 7, SEQ ID NO: 9 and
SEQ ID NO: 27.

Additionally, R_A or R_C may comprise one or more component sequences with a length of at
least 50 nucleotide residues and at least 90% identical to an aligned component sequence
of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID
NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

The nucleic acid molecules according to the present invention exist either in the form of
DNA or as RNA. Preferred embodiments are genomic DNA, cDNA, plasmid DNA or RNA
transcribed therefrom.

Nucleotides 437-2383 of SEQ ID NO: 1 encode a putative open reading frame of 648 amino
acids (SEQ ID NO: 10) which in SEQ ID NO: 1 is interrupted by a stop codon spanning
nucleotides 1631-1633. Nucleic acids encoding a protein comprising a component
sequence of at least 200 amino acids length being at least 85% identical to an aligned
component sequence of SEQ ID NO: 10 are a further preferred embodiment of the present
invention.

The nucleic acids according to the present invention represent an endogenous target of the transcriptional silencing system. Example 1 describes the cloning of specific embodiments of the present invention from Arabidopsis. The preferred size of transcribed RNA is between 1000 and 6000 nucleotides, particularly transcripts of about 1250, 2500, 4700 and 5000 nucleotides, which can be polyadenylated or not. The transcriptionally silent information present in the genome of wild type plants is found to be only expressed in a range of mutants affected in the maintenance of transcriptional silencing. Importantly, not only strains affected in transcriptional silencing through alterations of genome-wide DNA methylation, but also silencing mutants with unchanged methylation levels which do not show striking phenotypic alterations activate TSI, indicating that the release of silencing from endogenous templates does not require loss of methylation.

Initially two independent clones representing RNA which is specifically expressed in silencing mutants have been cloned. Anticipating that in wild type plants there are probably many more DNA templates suppressed by the silencing system, it is remarkable that parts of the two cDNAs cloned are closely related to each other and it is most likely that they are parts of the same transcript. The three main TSI transcripts of 5000, 2500 and 1250 nucleotides all contain a middle element isolated as TSI-A (SEQ ID NO: 5). The 5000 nt and the 2500 nt transcripts additionally enclose the second isolated element TSI-B (SEQ ID NO: 6), which is like TSI-A without protein coding capacity. The 5000 nucleotide long transcript further comprises a 5' extension (SEQ ID NO: 1 which is similar to SEQ ID NO: 2) encoding a putative open reading frame of 648 amino acids (SEQ ID NO: 10). The two 3' extension clones of TSI-A (SEQ ID NO: 3 and SEQ ID NO: 4) contain a region which can be aligned with nucleotides 1-569 of SEQ ID NO: 6 (nucleotides 808-1397 of SEQ ID NO: 3 and nucleotides 819-1365 of SEQ ID NO: 4) closely related to TSI-B (77 % identity).

Both the 5000 and the 2500 nucleotide transcripts are polyadenylated, while the most abundant transcript of 1250 nucleotides is absent from the polyA fraction of *mom1* RNA and might be retained in the nucleus.

All RNA species originate from unidirectional transcription, but it is not clear if they represent separate transcriptional units regulated by different promoters or if they are processing products of the same long transcript. A refined analysis of the TSI expression pattern is complicated by the multiplicity of potential chromosomal templates and their location mainly in the pericentromeric areas.

The novel TSI sequences do not reveal any putative function by sequence similarity to protein- or RNA-coding sequences. The only extensive similarity found was to the 3' half of the putative, degenerated retrotransposon *Athila* (Pélissier et al. 1995). The other part of *Athila* directly adjacent to the TSI template region was not reactivated in the silencing mutant. This suggests that epigenetic transcriptional silencing in *Arabidopsis* is not directed towards retrotransposons in general, although its targets may have originated from transposition events. This is further supported by the lack of transcriptional reactivation of other *Arabidopsis* retroelements, e.g. the *Ta* superfamily (Konieczny et al. 1991). Therefore, only specific pericentromeric repeats seem to be under epigenetic control, in the same way that only a subset of transgenic loci is susceptible to silencing. The existence of remnants of transposons is probably due to their chromosomal location rather than to sequence specificity, since degenerated retroelements have repeatedly been found in centromeric locations in fungi and plants.

One of the features proposed as a prerequisite for centromere function is late replication of the heterochromatic centromeres and pericentromeric areas in *Schizosaccharomyces pombe* and higher eukaryotes. If this was also true for *Arabidopsis* centromeres, undue loosening of suppressive chromatin leading to TSI expression could cause disturbances in mitosis, which would result in severe phenotypes. However, the *mom* mutant plants exhibit no abnormalities suggesting mitotic disorders. Therefore, transcriptional reactivation of some usually silent pericentromeric repeats, such as described here, does not impair their putative function. Alternatively, their silencing may be important under a specific, still undefined condition or on a longer time scale.

Finally, TSI expression is observed in cells growing for a long time in suspension culture. No release of TSI silencing is observed in any tissue of *Arabidopsis* wild type plants, including freshly initiated callus cultures. This suggests that an escape from the silencing control is not correlated primarily with dedifferentiation but could be the result of prolonged selection for fast growing dedifferentiated cells. Such a loss of silencing control could also underlie the accumulation of somaclonal variation during prolonged culture and resembles the situation in cells of actively proliferating carcinomas.

Nucleic acids according to the present invention are particularly useful in selecting plants which compared to wild type Arabidopsis plants of all available Arabidopsis ecotypes are

impaired in transcriptional gene silencing. A method allowing to select such plants comprises

a) separately preparing RNA of a series of plants;

b) probing said RNA preparations with a nucleic acid according to the present invention; and

c) identifying a plant whose RNA hybridizes with said nucleic acid.

In a preferred embodiment the probing step is performed after size fractionation of the RNA preparation by gel electrophoresis. For detection the probe is either radioactively labeled or labeled by other chemical modifications.

In another preferred embodiment of said method the step of probing consists of hybridizing the RNA with an oligonucleotide primer, extending said primer by reverse transcription and subsequent PCR amplification of the DNA generated using oligonucleotide primers specific for SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, or SEQ ID NO: 27. Plants which allow for the amplification of DNA fragments flanked by the oligonucleotide primers are identified as plants whose RNA hybridizes to the nucleic acid according to the invention.

Having available nucleotide sequence information of a genomic region, which is not expressed, i.e. transcriptionally silenced, in a wild type plant, allows to produce DNA representing at least part of a gene necessary to maintain silencing of this genomic region. Preferably the complete gene is produced. A corresponding method of production comprises

(a) mutagenizing wild type cells or plants by randomly inserting into their genomes a DNA tag with known sequence;

(b) identifying mutants of said cells or plants which express RNA that is not expressed in wild type cells or plants;

(c) cloning genomic DNA surrounding or close to the insertion site of the DNA tag;

(d) screening a genomic library of wild type cells or plants with the piece of genomic DNA obtained in process step (c) or a part thereof;

(e) identifying clones comprising at least part of the gene affected by the insertion of the DNA tag; and

(f) further processing the clones obtained in step (e) using recombinant DNA techniques.

In plant cells and plants mutagenesis is preferably achieved performing T-DNA insertion mutagenesis (Dilkes and Feldmann, 1998), or transposon tagging using the En/I (Pereira and Aarts, 1998) or the Ac/Ds system (Long and Coupland, 1998) as described in Arabidopsis protocols edited by Martinez-Zapater and Salinas, 1998. Other known physical or chemical methods of mutagenesis such as fast neutron irradiation or EMS mutagenesis (Feldmann et al., 1994) might require adaptation of the above method, but can be used for the production of equivalent DNA involved in the maintenance of silencing as well.

A convenient way to identify RNA that is expressed in mutant cells or plants but not in wild type cells or plants is reverse transcription of said RNA and subsequent PCR amplification of the generated DNA using oligonucleotide primers specific for said DNA (RT-PCR). This allows to pool the RNA of upto 1000 mutants which increases the speed of the identification step considerably.

The methods described above can be further elaborated and developed into a kit for the identification of plants impaired in transcriptional gene silencing. Such a kit necessarily comprises

a)   a nucleic acid according to the present invention conveniently labeled to be used as a hybridization probe or

b)   an oligonucleotide primer for reverse transcription of RNA and an oligonucleotide primer specific for a nucleic acid according to the present invention.

The oligonucleotide primer for reverse transcription can be a poly T primer or an oligonucleotide primer specific for a nucleic acid according to the present invention. The primers specific for nucleic acids according to the present invention are designed to allow PCR amplification of DNA templates characterized by the nucleotide sequences disclosed in the present invention.

EXAMPLES

Example 1:   Differential mRNA screening and cloning of *Arabidopsis* TSI sequences

Total RNA of the mutant line *mom1* (Amedeo et al, 2000) and its parental line A is isolated according to Goodall et al. (1990) using 2 g fresh weight of two-week-old seedlings. Polyadenylated RNA is obtained using Dynabeads Oligo (dT)$_{25}$ (Dynal). 2 µg of poly(A) RNA is used for suppression subtractive hybridization (SSH, Diatchenko et al, 1996) using the PCR-Select cDNA subtraction kit (Clontech) according to the suppliers' instructions. cDNA

derived from the mutant line *mom1* is used as tester and cDNA derived from the parental line A as driver cDNA population. The subtracted library is cloned into vector pCR2.1 (Invitrogen). 500 individual bacterial cultures from this library are grown according to the manual of the PCR-Select differential screening kit (Clontech). To reduce the number of false positive clones, the library is primarily screened by inverted Northern blots as described by von Stein et al. (1997). Twelve among the 500 primarily selected cDNA clones show increased abundance upon hybridization with labeled *mom1* cDNA. Direct Northern blot analysis comparing total RNA of the wild type and the mutant line with these 12 cDNAs as probes reveal a striking genotype-dependent differential expression for two of them. Said clones are sequenced using conventional rhodamine or dRhodamine dye terminators from PE Applied Biosystems and a Perkin-Elmer GeneAmp PCR system 2400, 9600 or 9700 thermocycler. The sequence reactions are analyzed using an ABI PRISM 377 DNA sequencer. The cDNA clones are named TSI-A (903 bp, SEQ ID NO: 5) and TSI-B (614 bp, SEQ ID NO: 6). Both are abundant in the *mom1* RNA but are undetectable in Arabidopsis line A and wild type Arabidopsis. No consistent differential expression between mutant and wild type is observed for the remaining 10 cDNAs.

5' and 3' extension reactions are performed using Clontech's Marathon Kit according to the manufacturer's instructions. Sequence specific primers are

TA-F1:   5'-TGGTTCACCAGATAAGCTCAGTGCCCTC-3' (SEQ ID NO: 11) and

TA-F2:   5'-CTTCAGACTGGATAGGACTAGGTGGGCG-3' (SEQ ID NO: 12, nested primer), for the 3'extension reaction and

TA-R1:   5'-CGCCCACCTAGTCCTATCCAGTCTGAAG-3' (SEQ ID NO: 13) and

TA-R2:   5'-CGCATCAAACAACTAACAACGAGGGCAC-3' (SEQ ID NO: 14, nested primer). for the 5'extension.

PCR amplification products are cloned into vector pCR2.1 (Invitrogen). Individual bacterial cultures are grown and subjected to colony PCR as described in the manual of Clontech's PCR-Select Differential Screening Kit, with the primer combinations used to create the extension reactions (Marathon Adapter primer Ap1 (Clonetech) combined with TA-R2 or TA-F2 for the 5'- or 3' extension reactions, respectively). To screen for positive TSI-A extension clones, the PCR products are blotted and hybridized to TSI-A. All PCR reactions for cloning procedures are performed with a polymerase mix performing proofreading activity (Advantage cDNA PCR Kit, Clontech).

Since only two transcripts are detected in the polyadenylated RNA fraction of *mom1* plants, this RNA is used for 5' and 3' extensions reactions starting from the TSI-A sequence. Two clones each are analyzed at the nucleotide sequence level. The 5' extension yields inserts of 2512 bp (clone a, SEQ ID NO: 1) and 1997 bp (clone b, SEQ ID NO: 2), which after alignment are 97 % identical to each other. The clones from the 3' extension have a length of 1682 bp (clone c, SEQ ID NO: 3) and 1652 bp (clone d, SEQ ID NO: 4) and are 94 % identical. Interestingly, both 3' extension clones of TSI-A contain a region of 569 bp closely related to TSI-B (77 % identity). This explains the detection of similar RNA species on Northern blots with TSI-A and TSI-B as probes and their hybridization to the same YAC and BAC clones, and suggests that TSI-A and TSI-B are part of the same polyadenylated transcript species expressed in the *mom1* mutant. To confirm that the 5' extensions of TSI-A are indeed part of the TSI transcripts, a *mom1* Northern blot is probed with a cDNA fragment close to the 5' end of the extension (probe ORF corresponding to nucleotides 943-1334 of SEQ ID NO: 1). Interestingly, only the about 5000 nt long transcripts in the poly(A) fraction hybridize to this probe. Since this class of transcripts hybridizes to TSI-A and TSI-B, the 5000 nt transcripts are probably produced from templates containing a particular order of all three sequence elements.

Example 2:   **Northern and Southern blot analysis and library screens**
Total RNA is either isolated as described by Goodall et al. (1990) or by the RNeasy Plant Mini Kit (Qiagen) according to the suppliers' instructions. For Northern blot analysis, the RNA is electrophoretically separated after denaturation by glyoxal in a 1.5 % agarose gel in phosphate buffer (pH 7) and blotted to nylon membranes (Hybond N, Amersham) using standard protocols. The Boehringer molecular weight marker I is used as a size standard. For Southern blot analysis, genomic DNA is isolated according to Dellaporta et al. (1983) and separated electrophoretically after endonucleolytic digestion. DNA fragments are transferred to nylon membranes (Hybond N, Amersham) according to standard procedures. Hybridization and washing of Northern and Southern blots and the filters with the YAC library is performed according to Church and Gilbert (1984). Probes are labeled with [α-^{32}P]-dATP by random prime DNA polymerization (Feinberg and Vogelstein, 1983) and exposed to X-ray sensitive film (Kodak X-OMAT AR).

Hybridization of <u>Northern blots</u> using total RNA prepared from 2-week-old seedlings with TSI-A of *mom1* visualizes four major transcripts with sizes of approximately 5000, 4700,

2500 and 1250 nucleotides. A TSI-B probe detects mainly two transcripts of 5000 nt and 2500 nt. Interestingly, the polyadenylated fraction of *mom1* RNA contains only the transcripts of 5000 nt and 2500 nt hybridizing to both cDNA probes (TSI-A and TSI-B). From the sizes of TSI transcripts detected it is obvious that both TSI clones represent only partial cDNAs. TSI expression is meiotically heritable and persists through 6 selfed generations of *mom1* with the same pattern of transcripts.

To examine TSI expression in other genotypes known to affect gene silencing, total RNA of several Arabidopsis mutant and transgenic strains known to be affected in gene silencing is probed with TSI-A. All the *som* mutants *som1* to *som8*, described by Mittelsten Scheid et al (1998) to be impaired in the maintenance of transcriptional silencing similar to *mom1*, show a high level of TSI-A expression. The mutation *ddm1*, originally identified to have decreased DNA methylation (Vongs et al. 1993), and later revealed to release transcriptional gene silencing from different loci (Mittelsten Scheid et al, 1998; Jeddeloh et al., 1998) also shows a high level of TSI-A expression. TSI-A expression is also expressed in a transgenic line described by Finnegan et al (1996) which shows decreased DNA methylation due to overexpression of DNA methyltransferase antisense mRNA as well as in a further Arabidopsis mutant affected in the DNA methyltransferase gene (said mutant is referred to as *ddm2* and has been provided by Eric Richards). Moreover, the silencing mutants *hog1* and *sil1*, but not *sil2* described by Furner et al (1998) express sequences hybridizing to TSI-A. Importantly, mutations affecting posttranscriptional gene silencing such as *sgs1* and *sgs2* described by Elmayan et al (1998) and egs1 described by Dehio and Schell (1994) do not express RNA which hybridizes to TSI-A.

Comparison of patterns of TSI-A expression in the different genotypes reveals genotype specific differences in the stochiometry of the different RNA species. *mom1* plants reveal a different expression pattern of TSI-A and TSI-B as compared to *som1* plants. These results indicate that a particular genetic deficiency in the transcriptional silencing system leads to a differential but specific activation of TSI templates. However, we observe variation in these activation patterns between different sources of plant material and different RNA preparations. Therefore, it is possible that patterns of TSI expression are more flexible and probably also controlled by still unknown factors acting in the mutant background, or by different stabilities among the transcript populations.

TSI-A and TSI-B are used as probes for Southern blots to determine the source of TSI transcripts and the organization of their template(s). The blots reveal that multiple copies of

TSI-A- and TSI-B-homologous sequences are present in the genome of Arabidopsis. Copy numbers are assessed by reconstruction experiments to approximately 130-300 copies of TSI-A.

To examine the degree of evolutionary conservation of the TSI arrangement, DNA of five Arabidopsis ecotypes (Zürich, Columbia, Landsberg erecta, Wassilevskija, C24) is compared by Southern blot analysis and hybridization to the TSI-A probe. Genomic DNA is digested with *Dra*I that has a single recognition site within TSI-A and *Ssp*I that does not cut within TSI-A. A significant conservation of the TSI-A pattern among different ecotypes is observed, with two main *Dra*I repeats of 4 kb and 1.3 kb and two major *Ssp*I fragments of 11 kb and 4 kb. Some minor differences specific for a particular ecotype indicate a limited genetic polymorphism within TSI-A. Probing the same membrane with TSI-B reveals complex banding patterns different in each ecotype which might indicate a lower degree of conservation for TSI-B, although the differences of the Southern blot patterns between TSI-A and TSI-B can also be explained if TSI-A is an internal part of a longer repeated element, and TSI-B is located proximal to a flank between repeated elements and variable single copy DNA regions.

After hybridizing TSI-A and TSI-B to the <u>CIC YAC library</u> covering 4 genome equivalents and 92% of the *Arabidopsis* genome sixty-two CIC clones out of 1152 turn out to hybridize with the TSI-A probe. Twenty-six of these contain also the pericentromeric 180-bp-repeat, 7 contain 5S RNA genes known to be located in the vicinity of a centromere, and 16 clones contain other markers that map close to centromeres. Only 4 of these clones map outside of centromeric regions. Similar mapping of TSI-B results in hybridization to all TSI-A positive CIC clones, with additional 7 clones hybridizing to TSI-B only. Thus both TSI repeats are concentrated in the pericentromeric regions of *Arabidopsis* chromosomes.

After hybridizing TSI-A and TSI-B to a cDNA library prepared with RNA isolated from *mom1* mutant plants, 22 hybridizing clones are further analyzed by sequence analysis. RNA is isolated from 2-week-old seedlings of the *mom1* mutant plant according to Goodall et al. (1990). The cDNA library is prepared using the Uni-ZAP XR library construction kit (Stratagene) according to the manufacturer's protocol. cDNA fragments larger than 500 bp are selected using the cDNA size fractionation columns from Gibco. 7 clones contain SEQ ID NO: 7 (TSI-A-15), 5 clones contain SEQ ID NO: 8 (TSI-A-2) part of which is identical to SEQ ID NO: 4.

- 11 -

Example 3:   **Database searches**

Sequence analysis is performed using the GCG software (Wisconsin Package Version10.0, Genetics Computer Group (GCG), Madison, Wisconsin). For identity searches, GenEMBL, the Arabidopsis thaliana database (http://genome-www3.stanford.edu/), Kazusa Arabidopsis opening site (KAOS; http://zearth.kazusa.or.jp/arabi/) and Swissprot are used. Peptide sequences are analysed by GeneQuiz (http://columba.ebi.ac.uk:8765/gqsrv/submit) or Expasy (http://www.expasy.ch/).

Searches within the genomic sequence databases GenEMBL, Arabidopsis thaliana Database, and KAOS confirm the presence of multicopy sequences related to TSI-A and TSI-B, which are distributed over all five chromosomes of *Arabidopsis thaliana*. Importantly, very often single BAC clones contain sequences homologous to both cDNA clones. In some cases, TSI-A and TSI-B related sequences are found more than once on the same BAC clone, suggesting that TSI-A and TSI-B belong to a clustered repetitive element.

The significant sequence heterogeneity between the cDNA classes and duplicates of the 5' and 3' extensions of TSI-A indicate that they originate from different activated repeats. To facilitate the data base search for a possible genomic template of the 5000 nt transcript among the multiple related copies, the overlapping cDNA sequences are combined to form a continuous 4860 bp sequence of "virtual" cDNA (SEQ ID NO: 9) which is used to search the Arabidopsis genomic sequence databases. A particular BAC clone (TAMU BAC T6C20, accession number AC005898) has 91 % identity to the combined cDNA sequence. Further, the search uncovers a chromosomal DNA stretch (BAC F7N22, accession number AF058825) 99% identical to the abundant cDNA A-15 of Example 2 (SEQ ID NO: 7). The genomic sequence of the transcribed region 5' to the region defined by SEQ ID NO: 7 is given in SEQ ID NO: 27. It is identical to nucleotides 65081-68202 of BAC F7N22. Both sequences are located at the pericentromeric region of chromosome five. The TSI sequence defined by nucleotides 65080 to 70370 on BAC F7N22 is 54 % identical to the retrotransposon-like repeat named *Athila*. The identity of this sequence as a retrotransposon is deduced from *Arabidopsis* genome sequences around heterochromatic regions that are marked by the presence of 180 bp satellite repeats. The 10.5 kb sequence of *Athila* has several characteristics of a retroelement, like long terminal repeats (LTR), a polypurine track (PPT) and a primer binding site (PBS) for tRNA priming of the reverse transcriptase, but its open reading frames do not share homology with proteins known to be

involved in transposition.

The TSIs map to the 3' terminal part of *Athila*. TSI-A covers a part of the 3' non-coding region of the putative retrotransposon and TSI-B corresponds to the PPT and a part of the 3' LTR. The sequence of the 5' TSI-A extension encodes a possible open reading frame of 648 amino acids length (SEQ ID NO: 10) with 51 % identity to 604 amino acids of the ORF2 deduced for *Athila*. The sequence coding for this ORF is also present on the TAMU BAC T6C20, however, the ORFs encoded by the two cDNAs clone a (SEQ ID NO: 1) and clone b (SEQ ID NO: 2) and the BAC are interrupted by translational stop codons after 398 amino acids (clone a), 83 amino acids (clone b) and 46/465/496/499/549 amino acids respectively (BAC T6C20). The ORF2 sequence present on BAC F7N22 is highly degenerated by five deletions of 2-31 bp and five insertions of 3-10 bp. This further supports the assumption that this sequence is derived from a putative but degenerated retrotransposon. Data base searches for proteins similar to the potential product of the 648 amino acids ORF do not yield significantly similar polypeptides neither to proteins usually encoded by retroelements nor to any other known polypeptides.

Example 4:   **RNAse protection assays**

RNase protection assays are performed according to Goodall et al. (1990) with minor modifications. To assay the direction of TSI transcription, the pCR2.1 based plasmid containing the TSI-A insert is cut by EcoRI creating a fragment of 781 bp which is ligated into the vector pGEM-7Zf(+) (Promega). To map the 5' transcription start, the probe is generated by amplifying the BAC F7N22 region between positions 64929 and 65567 and inserting the product into the pGEM-7Zf(+) vector (Promega). Labeled probes are synthesized by *in vitro* transcription of the linearized plasmid in the presence of $[\alpha-{}^{32}P]$-UTP using T7 polymerase (Promega) or Sp6 polymerase (Boehringer) and purified by electrophoresis (Goodall et al., 1990). Single stranded RNA is cleaved by either 4 µg RNase A and 0.6 U RNase T_1 (RNase A/T assay) or by 20 U RNase T_1 (RNase T assay). Protected fragments are separated on a denaturing 6 % polyacrylamide gel. The dried gel is exposed to a PhosphorImager screen (Molecular Dynamics) and to X-ray sensitive film.

To determine the polarity of TSI transcription, RNase T and RNase A/T protection assays are performed with TSI-A probes of opposite polarity. TSI-A sequences are used as probes since TSI-A is present in all transcripts detected on Northern blots. There is no evidence for protection of the probe corresponding to the sense strand. This suggests the lack of any

TSI antisense RNA and a unidirectional transcription of the TSI templates. Interestingly, RNase digestions with a TSI-A antisense probe creates a complex pattern of TSI-A protected bands. This suggests that many different but related RNAs hybridize to the probe. Since a fragment of 781 nt as expected for the protection of the entire TSI-A probe is visible, it can be concluded that TSI-A is part of an activated transcript throughout and not an artifact generated by template switch during the SSH procedure. Furthermore, some of the protected TSI-A fragments are clearly more abundant than others, suggesting either a structural conservation of particular regions of TSI-A within related RNAs, or alternatively a higher abundance of certain transcript subspecies.

The sequence information of BAC F7N22 is used to determine the position of the transcription start for the longest TSI transcript. An antisense RNA probe for RNase A/T protection is produced spanning the 638 nucleotides between positions 64929 and 65567. The probe is hybridized with total RNA from *ddm1*, *som7* and *mom1*. In all RNA preparations, a fragment of approximately 480 nt (±10 nt) is protected and allowes positioning of the TSI transcription start on BAC F7N22 to 65087 (± 10 nt) in different mutants.

Example 5:   **Reverse Transcription PCR (RT PCR)**

Reverse transcription is performed with 1 μg total RNA from *mom1* in the presence of 1 mM dNTPs, 4-20 U RNasin (Promega), 1x AM RTase buffer (Boehringer) and 25 U AM reverse transcriptase (Boehringer) at 37°C for 1 hour, followed by heat inactivation of the reaction mixture. As template for PCR 50 ng reverse transcribed RNA primed by gene specific antisense primers (BA-R1, BA-R2, AT-R1, and TA-R1, see below) or 100 ng genomic DNA or 100 ng cDNA are used. PCR is started with 3 min denaturation at 94°C, followed by 30 amplification cycles (denaturation at 94°C/30 sec, annealing at 62°C/30 sec, and elongation at 72°C/30 sec) in the presence of 0.2 mM dNTPs, 0.4 μM forward and reverse primers, 1x Taq DNA polymerase buffer (Boehringer) and 0.25 U Taq DNA polymerase (Boehringer). The nucleotide sequences of the primers used for RT-PCR are:

AT-F1:   5'-CGATAACATCGACCGTATTGCTCGCC-3' (SEQ ID NO: 15)

AT-R1:   5'-AACTAGCTCCCATCCGTCTTCGACATCC-3' (SEQ ID NO: 16)

AT-F2:   5'-TGCATCACACCGGATTGGATTGAC-3' (SEQ ID NO: 17)

AT-R2:   5'-TGTTCCCCTGAACCATAGCAATGAGACC-3' (SEQ ID NO: 18)

BA-F1:   5'-CAAACAGACAGAGTGTGGCCCACCACC-3' (SEQ ID NO: 19)

BA-R1:   5'-AAGAGAGGGAGAAGGCAGTGGCGTGAG-3' (SEQ ID NO: 20)

BA-F2:   5'-TGCAAACCCACAGGACCAAGTCTACCC-3' (SEQ ID NO: 21)

BA-R2:   5'-ACAGATGGTGATAGCGTGAGCGGTGGC-3' (SEQ ID NO: 22)

F7-F1:   5'-TCAACCTTTTGCCCCAACAACCACTC-3' (SEQ ID NO: 23)

F7-R1:   5'-TCTCCATCCACGCTTTCCTGAATGTCC-3' (SEQ ID NO: 24)

GS-F1:   5'-GGAGAAGGAAGCTGAAAATCATATTGTGG-3' (SEQ ID NO: 25)

GS-R1:   5'-ATGATGATCCTAAGTCTACCCTTTTGCAC-3' (SEQ ID NO: 26)

As a positive control for the PCR reactions, the TA-F1 and TA-R1 primers are used. The
reverse transcriptase region of the pol gene of Arabidopsis Ty1/copia-like retrotransposon
family is amplified as described by Konieczny (1991) with immaterial modifications.

Since the nucleotide sequence of the TSI transcripts TSI-A and TSI-B is related to the
nucleotide sequence of the 3' half of retrotransposon-like element *Athila* including the
second ORF and the 3' LTR, we examined by RT-PCR whether transcription of the 5' part
of *Athila* including the first ORF is activated in *mom1* plants. Five primer pairs are chosen
(BA-F1 - BA-R1; BA-F2 - BA-R1; AT-F1 - AT-R1; AT-F2 - AT-R2, F7-F1 - F7R1) according
to the sequence information about *Athila* and the related parts of BAC T6C20 and BAC
F7N22. All primer combinations amplified the expected products from genomic template
DNA, but no PCR product could be obtained from *mom1* RNA, regardless, whether cDNA
synthesis was started from an *Athila-* or BAC-specific reverse primer or from polyT-primed
cDNA (data not shown). Activation of TSI therefore is limited to sequences related to the 3'
part of *Athila*.

The two classes of isolated cDNAs share only approximately 50% identity with *Athila*. To
directly address the question of whether *Athila* is expressed, RT-PCR experiments are
performed with *Athila*-specific primers (GS-F1, GS-R1) in the TSI homologous region.
However, the corresponding fragment cannot be amplified from RNA of *mom1* seedlings,
suggesting that only a subset of *Athila*-like sequences but not the *Athila* element itself is
reactivated in the mutant background.

To investigate whether other retroelements are transcriptionally activated in *mom1*,
degenerated primers in a conserved region of the reverse transcriptase gene used to clone

and describe the Ta superfamily of Arabidopsis retrotransposons (Konieczny et al., 1991) are used to investigate, if members of the Ta family are transcribed in the mutant background. Although the expected 268 bp fragment can be amplified from genomic DNA, no amplification is achieved in RT-PCR with *mom1* RNA as template. This indicates, that, in spite of the TSI homology to retrotransposons, these elements are not generally activated in the *mom1* mutant.

Example 6:   **TSI expression after application of stress (salinity, UV-C, pathogen)**

Induction of TSI upon <u>UV-C</u> is tested on Northern blots with RNA samples from 1 week-old-seedlings subjected to UV-C treatment of 1 kJ/m^2 or 5 kJ/m^2 which are collected at several time points within 1 hour (Revenkova et al., 1999). The effect of osmotic stress is tested on Northern blots with RNA from one-week-old seedlings that are transferred for 24 hours to medium with NaCl concentrations of 0, 0.04, 0.08 and 0.12 M (Albinsky et al., 1998). To test TSI expression upon <u>pathogen</u> stress, RNA of 3-week-old seedlings either mock treated or infected with *Peronospora* is analysed by Northern blot analysis. To verify the appropriate pathogen response, induction of PR1 expression is monitored by reprobing the membrane with a PR1 probe.

In young seedlings (2 weeks old) and in different tissues of mature wild type plants (roots, shoots, leaves, flowers, siliques), TSI expression cannot be detected. The application of various stress treatments namely elevated salinity, UV-C, or pathogen infection, does not activate TSI in wild type plants. TSI expression is also not detected in freshly initiated callus cultures, and transcriptional suppression of TSI is stable even after several *in vitro* passages of the callus culture. However, the only exception so far are cells derived from wild type *Arabidopsis* (literature) growing for a long time in suspension culture. These cells express TSI-A, indicating release of TSI silencing under these conditions.

REFERENCES

- Albinsky et al, *Plant J.* **17**: 73-82, 1999

- Amedeo et al, *Nature* **405**: 203-206, 2000

- Church and Gilbert, *Proc Natl Acad Sci USA* **81**: 1991-1995, 1984

- Dehio and Schell, *Proc Natl Acad Sci USA* **91**: 5538-5542, 1994

- Dellaporta et al, *Plant Mol Biol Rep.* **1**: 19-21, 1983

- Diatchenko et al, *Proc Natl Acad Sci USA* **93**: 6025-6030, 1996

- Dilkes and Feldmann, *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 339-351, 1998

- Elmayan et al, *Plant Cell* **10**: 1747- 1757, 1998

- Feinberg et al, *Anal Biochem.* **132**: 6-13, 1983

- Feldmann et al, *in* Arabidopsis (Ed. Sommerville and Meyerowitz), CSHL Press, New York, pp 137-172, 1994

- Finnegan et al, *Proc Natl Acad Sci USA* **93**: 8449-8454, 1996

- Furner et al, *Genetics* **149**: 651-662, 1998

- Goodall et al, *Methods Enzymol.* **181**: 148-161, 1990

- Jeddeloh et al, *Genes Dev.* **12**: 1714-1725, 1998

- Konieczny et al, *Genetics* **127**: 801-809, 1991

- Long and Coupland *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 315-328, 1998

- Mittelsten Scheid et al, *Proc Natl Acad Sci USA.* **95**: 632-637, 1998

- Pellisier et al, Plant Mol Biol. **29**: 441-452, 1995

- Pereira and Aarts, *in* Arabidopsis protocols, Methods in Molecular Biology **82** (Ed. Martinez-Zapater and Salinas), Humana Press, Totowa, New Jersey, pp 329-338, 1998

- Revenkova et al, *EMBO J.* **18**: 490-499, 1999

- von Stein et al, *Nucleic Acids Res.* **25**: 2598-2602, 1997

- Vongs et al, *Science.* **260**: 1926-1928, 1993

What is claimed is:

1. A nucleic acid having the formula R_A-R_B-R_C, wherein
    -- R_A, R_B and R_C constitute component sequences consisting of nucleotide residues independently selected from the group of G, A, T and C or G, A, U and C, wherein
        G is Guanosinmonophosphate,
        A is Adenosinemonophosphate,
        T is Thymidinmonophosphate,
        U is Uridinmonophosphate and
        C is Cytidinmonophosphate;
    -- R_A and R_C consist independently of 0 to 6000 nucleotide residues;
    -- R_B consists of at least 50 nucleotide residues; and
    -- the component sequence R_B is at least 80% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

2. The nucleic acid of claim 1, wherein R_B consists of at least 100 nucleotide residues and is at least 85% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

3. The nucleic acid of claim 1, wherein the nucleic acid sequence of R_B is the sequence given in SEQ ID NO: 7, SEQ ID NO: 9 or SEQ ID NO: 27.

4. The nucleic acid of claim 1, wherein R_A or R_C comprise one or more additional component sequences with a length of at least 50 nucleotide residues and at least 90% identical to an aligned component sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9 or SEQ ID NO: 27.

5. The nucleic acid of any one of claims 1 to 4 comprising an open reading frame encoding a protein comprising a component sequence of at least 200 amino acids length being at least 85% identical to an aligned component sequence of SEQ ID NO: 10.

6. A method of selecting a plant which compared to a wild type plant is impaired in transcriptional gene silencing, comprising
    a) separately preparing RNA of a series of plants;
    b) probing said RNA preparations with a nucleic acid according to claim 1; and
    c) identifying a plant whose RNA hybridizes with said nucleic acid.

7. The method of claim 6, wherein process steps b) and c) comprise reverse transcription of the RNA and subsequent amplification of the generated DNA using oligonucleotide primers specific for SEQ ID NO: 1, SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 9 or SEQ ID NO: 27.

8. A method of producing DNA representing at least part of a gene necessary to maintain silencing of another gene in a cell or plant, comprising
    (a) mutagenizing wild type cells or plants by randomly inserting into their genomes a DNA tag with known sequence;
    (b) identifying mutants of said cells or plants which express RNA that is not expressed in wild type cells or plants;
    (c) cloning a piece of genomic DNA surrounding or close to the insertion site of the DNA tag;
    (d) screening a genomic library of wild type cells or plants with the piece of genomic DNA obtained in process step (c) or a part thereof;
    (e) identifying clones comprising at least part of the gene affected by the insertion of the DNA tag; and
    (f) further processing the clones obtained in step (e) using recombinant DNA techniques.

9. The method of claim 8, wherein plants are mutagenized using T-DNA, Ac/Ds or EN/I insertions, chemical or physical mutagenesis.

10. The method of claim 8, wherein mutants expressing RNA that is not expressed in wild type cells or plants are identified by reverse transcription of said RNA and subsequent amplification of the generated DNA using oligonucleotide primers specific for said DNA (RT-PCR).

11. The method according to claim 10, wherein pools of mutants are analyzed by RT-PCR.

12. The method according to claim 8, wherein the complete gene necessary to maintain silencing is produced.

13. A kit for the identification of plants impaired in transcriptional gene silencing comprising
    a) a nucleic acid according to claim 1 conveniently labeled to be used as a hybridization probe or
    b) an oligonucleotide primer for reverse transcription of RNA and an oligonucleotide primer specific for a nucleic acid according to claim 1.

(51) International Patent Classification[7]: **C12N 15/82**, C07K 14/415, C12Q 1/68, A01H 5/00

(21) International Application Number: PCT/EP00/08994

(22) International Filing Date:
14 September 2000 (14.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
9921964.4   16 September 1999 (16.09.1999)   GB

(71) Applicant *(for all designated States except US)*: **SYNGENTA PARTICIPATIONS AG** [CH/CH]; CH-4058 Basel (CH).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **STEIMER, Andrea** [CH/CH]; Solothurnerstrasse 92, CH-4053 Basel (CH). **MITTELSTEN SCHEID, Ortrun** [DE/CH]; Martinskirchplatz 15, CH-4051 Basel (CH). **PASZKOWSKI, Jerzy** [PL/CH]; Blauenweg 10, CH-4224 Nenzlingen (CH).

(74) Agent: **SYNGENTA PARTICIPATIONS AG**; Patent and Trademark Dept., Site Rosental, CH-4002 Basel (CH).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSCRIPTIONALLY SILENCED PLANT GENES

(57) **Abstract:** The invention relates to gene silencing as observed after integration of transgenes into plant genomes. Comparison of transcriptional gene expression between an Arabidopsis line carrying a silent transgene present in multiple copies and its mutant derivative *mom1* impaired in silencing of the transgene revealed two cDNA clones which are expressed in the mutant plants, but not in the parental and not in wild type plants. Both clones are derived from the same family of transcripts referred to as TSI (Transcriptionally Silent Information). Genomic templates encoding TSI are repetitive elements with mainly pericentromeric location and conserved organization among various ecotypes. Transcriptional silencing of the genomic TSI templates is specifically released in the mutant Transcription of TSI can be used as a marker to identify a defective silencing pathway in a plant.

ET 327548479 US

DECLARATION AND POWER OF ATTORNEY FOR UNITED STATES PATENT APPLICATION

☒ Original   ☐ Supplemental   ☐ Substitute

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name, and

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if more than one name is listed below) of the subject matter which is claimed and for which a United States patent is sought on the invention entitled

TRANSCRIPTIONALLY SILENCED PLANT GENES

the specification of which:

☐   is attached hereto.

☐   was filed on _____ as Application No. _____
               (day/month/year)

and, if this box (☐) contains an ✕

☐   was amended on _____
                   (day/month/year)

☒   was filed as Patent Cooperation Treaty international Application No.

PCT/EP00/08994 _____ on ____**14/09/2000**____
                                      (day/month/year)

and, if this box (☐) contains an ✕

☐   entered the national stage in the United States and was accorded Application No.

_____

and, if this box (☐) contains an ✕

☐   was amended, subsequent to entry into the national stage, on _____
                                                                              (day/month/year)

I hereby state that I have reviewed and understand the contents of the above-identified specification, including the claims, as amended by any amendment(s) referred to above and, if this application was filed as a Patent Cooperation Treaty international application, by any amendments made during the international stage (including any made under Patent Cooperation Treaty Rule 91, Article 19 and Article 34).

I acknowledge my duty to disclose all information which is known by me to be material to the patentability of this application as defined in 37 C.F.R. § 1.56.

I hereby claim the benefit under 35 U.S.C. §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate listed below and under 35 U.S.C. §365(a) of any Patent Cooperation Treaty international application(s) designating at least one country other than the United States listed below and have also listed below any foreign application(s) for patent or inventor's certificate and Patent Cooperation Treaty international application(s) designating at least one country other than the United States for the same subject matter and having a filing date before that of the application the priority of which is claimed for that subject matter:

COUNTRY/REGION (OR P.C.T.)	APPLICATION No.	FILING DATE (day/month/year)	PRIORITY CLAIMED
GB	9921964.4	16/09/1999	☒ Yes ☐ No
			☐ Yes ☐ No
			☐ Yes ☐ No
			☐ Yes ☐ No
			☐ Yes ☐ No

I hereby claim the benefit under 35 U.S.C. § 119 (e) of any United States provisional application(s) listed below:

APPLICATION NO.	FILING DATE (day/month/year)

I hereby claim the benefit under 35 U.S.C. §120 of any United States application(s) listed below and under 35 U.S.C. §365(c) of any Patent Cooperation Treaty international application(s) designating the United States listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in said prior application(s) in the manner required by the first paragraph of 35 U.S.C. §112, I acknowledge my duty to disclose all information known to me to be material to patentability as defined in 37 C.F.R. §1.56 which became available between the filing date(s) of the prior application(s) and the national or Patent Cooperation Treaty international filing date of this application:

United States Application No.	United States Filing Date (day/month/year)	Status (Pending, Abandoned or U.S. Patent No.)	International Application No. and Filing Date (day/month/year)

I hereby appoint the registered practitioners associated with Customer No. 022847, respectively and individually, as my attorneys and agents, with full power of substitution and revocation, to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith.

If these brackets contain an X [X], I hereby authorize the registered practitioners associated with Customer No. 022847 and any others acting on my behalf to take any action relating to this application based on communications from the Patents and Trademarks Division of Syngenta Crop Protection AG, Basle, Switzerland, or an affiliate thereof or a successor thereto, without direct communication from me.

Please address all communications to the address associated with Customer No. 022847, which is currently Larry W. Stults, Syngenta, Patent and Trademark Department, 3054 Cornwallis Road, Research Triangle Park, NC 27709.

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. §1001 and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Full name of sole or first joint inventor **Andrea STEIMER**

Inventor's signature _Andrea Steimer_   Date 06/02/2002

(day/month/year)

Residence **8057 Zürich, Switzerland**

Citizenship **Swiss**

Post Office Address **Langackerstrasse 49**
8057 Zürich
Switzerland

IMPORTANT: Before this declaration is signed, the patent application (the specification, the claims and this declaration) must be read and understood by each person signing it, and no changes may be made in the application after this declaration has been signed.

US   Case   S -31147A

| Full name of second joint inventor, if any | **Ortrun MITTELSTEN SCHEID** |

Inventor's signature _____ Date __14/ 02 / 2002__
(day/month/year)

| Residence | **4051 Basel, Switzerland**   SEX |

| Citizenship | **German** |

| Post Office Address | **Martinskirchplatz 15**
4051 Basel
Switzerland |

| Full name of third joint inventor, if any | **Jerzy PASZKOWSKI** |

Inventor's signature _____ Date __14 02. 2002.__
(day/month/year)

| Residence | **4224 Nenzlingen, Switzerland**   SEX |

| Citizenship | **Polish** |

| Post Office Address | **Blauenweg 10**
4224 Nenzlingen
Switzerland |

| Full name of fourth joint inventor, if any | |

Inventor's signature _____ Date _____
(day/month/year)

Residence

Citizenship

Post Office Address

SEQUENCE LISTING

<110> Novartis AG
      Novartis Research Foundation

<120> Transcriptionally Silenced Plant Genes

<130> S-31147A

<140>
<141>

<150> GB 9921964.4
<151> 1999-09-16

<160> 26

<170> PatentIn Ver. 2.1

<210> 1
<211> 2512
<212> DNA
<213> Arabidopsis thaliana

<400> 1
```
tcgccgttcc agcttcgcat actctctcac cgtctctcgt ttcactcgac cacttcacac 60
ttcgcctcaa catcttcgcc ggagtttctc gccattgtcc gtgcttccgt catctccgtt 120
cactcgacca ccggaccggc ttcaccatct ctcaactatc caccattcac tcgacctcgc 180
cattcactgc gcctccattc gtctctttac tcgactgctc ctcaaaccgc caccgtcttc 240
tctaaattcg ccgtttactc gaccacactg ttacgtctct cattcgtgta cagtcgaccg 300
ctatacccga agccacaata tcactctact cgaccgtttc actcgatcgc gtacttgact 360
ggtttagtgt gtgtgtttat ttgaactaac atattgatat ttggttttga gttacattct 420
ttttcaggga atcaatatga gcaactacag tggcgaatcc tccatggatg cggattacaa 480
cgtcgatgaa gctgaatctt ggtcaactag accagagaga gagcaacagg cttatgagag 540
cttcagagcc gagacccaac gctcagtagc tcgacgcaat gaaaggagag ctgagattgc 600
tagaggaaag agagcgatga ccagcagata tgagttgatc gacgaagata ttgacgtcga 660
gtatgagcct gagtcatggc acagagaaac aaaactgttg aacaagcctg atgaagttac 720
agtggaagag tacatcagac ttttcgagct gaacgacttc tggggagcga ggtacccctg 780
ttatgagact ctagcccagc ttaggctact ggaggacgta cagcacttat tcgagaagtg 840
ccatcttgag acgctgatgt cttacccgta cgtcgcttac aagaaggaaa caatagagtt 900
tctctccact ctgcaagtgg agttgtatca gggacttact gcagatgaac tggagagtga 960
agggttggga ttcttgactt tttcagttaa cgagcagcgt taccagctat ctatcaagag 1020
cttggaagga ttatttggtt ttcccagtgg aaagggaact aaacccaagt tcgaaaggga 1080
agagttgaag gatttgtggt taaccattgg gaacgatttg gcgctcaact ctgcaaggtc 1140
taagagcaac cagattcgaa gccctgtgat ccgctactat cagcgctcag tagcgaatgt 1200
tctgtacccc agggaatcta caggcaccgt gtctaacaca gacatggaga tgattgattc 1260
tgcactcaag ggtattctcc ggagaacaaa ggggaagaag gtcctaaagg gcgaccttaa 1320
tgatacacca ccggtcatgc ttctgttgat ccatatgtgt ggatacagga agtgggcgca 1380
caccaacggg aggaagaagg tgcgaggagc cctttgtgtg ggtggcgttg tgacaccgat 1440
tctgattgca tgtggtgtac ctctcacgtc tccagggttt gatccgagga tgatggattt 1500
agatcacttg cgtcgttgtg agttctgga gtacgacatg gttggcgatt tctatcgcta 1560
caaattcgag cactccctga cccgaacagc caacattttg cttccctgca tcgaggccac 1620
aaccatactt tagggtgaga acattgactt cagacctgcg cgtgattacc tctactttga 1680
gagcactcca ccgactgatg acaatgtccc tacgacggaa gctacagagg atgattttgc 1740
tgagacggat gaggataggg aggaggagta tgatacgagc atgtatcatt tcagtgagca 1800
cgtacctcca gcgcaggaga gcaagagctt gagtgaagct cacagaaaca acagtaagtt 1860
```

```
gcagaggtgg tgcaagaaac aagataggtt acttatcaag tgcttcaagg ccatcacgtt 1920
tctaacggac aagataagtt gcttctcttc taccacagct attccgcagg gagagcgtcc 1980
tcaggacatg ccttcgaaga gatatgacgc gccagggcca agtcatcaca ggcctgagcc 2040
aagtcaccac aggcctgagc ctagtgaccg agtagtacca ccagtccctg caaggcattc 2100
atcattcgag cctcgggagc tcgggagaaa gaagaaggct gcactcgcta ggtctggcag 2160
caggagtaga cgacttctcc agtcccgtag cttacgcgac cgcggtgctg ccgcagcag 2220
aagaagagag gtcgagtatc atcagagcgg tgctggccgc ggcgaaggag cagaggtcga 2280
gtaccccag ggggaagctg agacacaaca gggagattct tcgatggcct gggagcaatc 2340
acaggcagct attgacgacc aactccgctc cttcttccac tgaggtatgc acctcactcc 2400
accattgtaa tataccatct cttgttttt attttgtttt tgtgatgtgt tttgtcctga 2460
gtactctctt ccaaatttgg tcacacagtg gactgtgtga tttaagtttg gg        2512
```

```
<210> 2
<211> 1997
<212> DNA
<213> Arabidopsis thaliana

<400> 2
ccgggcaggt caacaggctt atgagagctt tagagctgag acccaacgct cagtagctcg 60
acgcaatgaa aggagagctg agattgctag aggaaagaga gcaatgacca gcagatatga 120
gttgatcgac gaagatattg acgtcgagta tgaacctgaa tcatggcaca gagaaacgaa 180
gctgttgaac aagcccgatg aagttacagt agaggagtac atcagacttt tcgagctgaa 240
cgacttctag ggaacgaggt acccctgtta tgagacttta gcccagctgg ggctactgga 300
ggacgtacat cacttattcg agaagtgcca tctggagacg ctgatgtctt acccgtacgt 360
cgcttacaag aaggaaacaa tagagtttct ctccactctg caagtggaga tgtatcaggg 420
acttactgca gatgagctgg agagtgaagg gttggggttc ttgactttt cagttaacga 480
gcagcgttac cagctatcta tcaagagctt ggaaggatta tttggttttc caagtggaaa 540
gggaactaaa cccaagttcg agagggaaga gttgaaagat ttgtggttaa ccattgggaa 600
cgatttggca ctcaactctg caaggtctaa gagcaaccag attcgaagcc ctgtgatccg 660
ctactatcag cgctcagtag cgaatgttct gtaccccagg gaatctacag gcatcgtgtc 720
taacacagac atggagatga ttgatgttgc actcaagggc attctccgga gaacaaaggg 780
gaagaaggtc ctaaagggcg accttaatga tacaccaccg gttatgcttc tgttgatcca 840
cctgtgtgga tacaggaagt gggcgcacac caacgagaag aagaaggtgc gaggagccct 900
ttgtgtaggt ggcgttgtga caccgattct gattgcatgt ggtgtacctc tcacgtctcc 960
agggtttgat ccgaggatga tggatttaga tcacttgcgt cgttgtgagt ttctagagta 1020
cgacatggtt ggcgatttct atcgctacaa attcgagcac tccctgaccc gaacagccaa 1080
cattttgctt ccctgcatcg aggccacaac catacttcag ggtgagaaca ttgacttcag 1140
acctgcgcgt gattacctct actttgagag cgctccaccg actgatgaca atgtccctac 1200
gacggaagtt acagaggatg atattgctga gacggatgag gataggggagg aggagtatga 1260
tacgagcatg tatcatttca gtgagcacgt acctccagcg cgggagagca agagcttgag 1320
tgaagctcac agaaacaaca gtaagttgca gaggtggtgc aagaaacaag ataggctact 1380
tatcaagtgc ttcaaagcca tcacgtttct gacggacaag ataagctgct tctcttctac 1440
cacagctatt ccgcagggag agcatcctca ggacatgcct tcaaggagat atgacgcgcc 1500
agcgccaagt catcacaggc ctgagccaag tcaccacagg cctgagccta gtgaccgagt 1560
agtcccacca gtccctgcaa ggcattcatc attcgagcct cgggagctcg ggagaaagaa 1620
gaaggctgca ctcgctcggt ctggcagcag gagtacacga cttctacagt cccgtagctt 1680
acgcgaccgt ggtgctggcc gcagcagaag aagagaggtc gagtatcatc agagcggtgc 1740
tggccgcgac gaaggagcag aggtcgagta ccccacgggg aagctgagac acaacaggga 1800
gattcttcga tggcctggga gcaatcacat gcagctattg acgaccaact ccgctccttc 1860
ttccactgag gtatgcacct cactccacca ttgtaatata ccatctcttg tttttatttt 1920
gttttgtga tgtgttttgt cctgagtact ctcttccaaa tttggtcaca cagtggactg 1980
tgtgatttaa gtttggg                                                1997
```

```
<210> 3
```

```
<211> 1682
<212> DNA
<213> Arabidopsis thaliana

<400> 3
cttatattat gggttgggat gtgtttaaag aaaaggggga attcattgtt gataaggaaa 60
gggaaagaat tctaggggaa gtaagctaaa gaagttagaa aaaatctagt aaaggttttg 120
ggaatgttaa agaaaagaat gaggttcttg ttagctaaag aataagggtt aaaagccttt 180
ggttttaaag attaaaaaaa aaacaggaac cttagttgtt aaagaaatcc aaacccgcta 240
gatgtatcaa gagcgttgag aaagcttctc ctagagttaa gagaaaagaa aagaatgata 300
tgaaaaagag tttgaaagat tcatgagtgc aaagggtaga gttaagttgg gacaggagtt 360
ggttttacca ttagaacttc attgttatac tctgggtaga tgggatctta tctctgtatg 420
cataatttgg gacttacctt tagcattcta ctaaagctca atcattcttg agggatcccc 480
tgttacttaa gcctattcta taagggacca tctttgtctc ttgaccttca ccttggccga 540
atgagttcat tgatgatgca ttgcttgatt cgcgttccag aactaatgaa tgttaaaggg 600
attggtagat ttgaaagcat gtgtaggtcg agtataagag acggattgat tgaaaacaag 660
gcatggctaa cgttttgag tagaattcaa tcatatcgca tcttagaact accaacttgg 720
acattgattt tatttgctct atcatatgct ttggttttga gtccccgcct tcactcctct 780
ccttcaacta tgtcttctta tttgcttgag ggcaagcaaa gactaagttt gagggagttg 840
atatgtctat aatttgcatg ttttcagtgt ccattcatca tcgttttgag tccagtttcg 900
tatcattcat cactgtttta tatcatttct catcattctt gcatactttg catgattagg 960
ataactttgc atacatattg catttctgag ttgttttcag gtgatttgga gctgtttgca 1020
agcaaattgg aagaaatgag ccagaaccag aagacatact cgacccctag gtcgagtgac 1080
tttggggcca ttcttcccac atactcggcc cccaggtcga gtgactttgg agccattctt 1140
cccatccact cgaccaccgg gtcgagtaac cttagctcag gccactcgat gacactactc 1200
gaccccaggt cgagtatcac ttcgccacac cacctgacaa cactcgacca atcactctac 1260
caagttactc gaccccctgg tcgagtatca tcactcacca ccatcagcat cactcgaccg 1320
gacactcgat cacgtcttca cagtctactc aaatccgcag tcaaccagac aagctgagca 1380
caaggaagag aagaggagaa gacaaagtgc ttggaagcgg cctggacctc catcggatca 1440
cgaagcccat ctcggcccat tatctctcta tgggccgagc gattaggtta ttggcccgtc 1500
tactatcatt ttatttcgtt ttgtataaat agatgtctta gggtttgtc ctgagacatc 1560
tagtcgacat tgagtttttt ttgcttcagt tttatttct gttctactct gctgcgccgc 1620
ttttgcttct gcaacctgta attcgagatt tttccaagtt attcagattc cgcatttgat 1680
tt                                                                1682

<210> 4
<211> 1652
<212> DNA
<213> Arabidopsis thaliana

<400> 4
cttatatcat gggtttggat cagtttaaaa aaaaaaaagg gtgaattcat tgttgataag 60
gaaagggaaa gaattctagg ggaagtaagc taagaagtt agaaaaaaaa aaatctagta 120
aaggttttgg gaatgttaaa gaaaagaatg aggttcttgt tagctaaaga agaagggtta 180
aaagcctttt gttttaaaga ttaaaaacag gaaccttagt tgttaaagaa atccaaatac 240
gctagatgta tcagagtgtt gagaaagctt ctcctagagt taagagaaaa gaaaagaatg 300
atatgaaaaa gagtttgaaa gattcatgag tgcaaagggt agagttaagt cttgtattg 360
ggactggagt tgggattacc attagagctt cattgttata ctatgggtag atgggatttt 420
atctctgtat gcataacttg ggacttacct ttagcattct actaaagctc aatcattctt 480
gagagatccc ctgttactta agcctattct gtaagggacc atctttgtct cttgaccttc 540
accttagcca aatgagttca ttgatgatgc attgtttgat tcacgttcca gaactaatga 600
atgttaaagg gattggtaga tttgaaagca tgtgtaggtc gagtataaga acggattga 660
ttgataacaa ggcatggcta acgttttcga gtaaaattca atcatatcgc atcttagaac 720
taccaacttg gacattgatt ttatttgctc tatcagatgc tttggttctg agtccccacc 780
ttcaaacctc tccttcaact atgtcttctt atttgcttga gggcaagcaa agactaagtt 840
```

```
tgggggagtt gatatgtcta taatttgcat gttttcagtg tccattcatc atcgttttga  900
gtccagtttc gtatcattca tcactgtttt atatcatttc tcatcattct tgcatacttt  960
gcatgattag gatagctttg tatacatatt gcatttctga gttgttttta ggtgatttgg 1020
agctgtttgc gagcaaattg gaagaaacga gccagaacaa gaagccatac tcgaccccct 1080
ggtcgagtga ctttggagcc attcttccca tctactcgac ccggggtcg agtaacctca 1140
gctcaggcca ctcgatgacg ccactcgtcc ccctggtcg agtatcactt cgccacacca 1200
cctgaccaca ctcggccgtt cactctacca cgttactcga cccctggtc gagtatcatc 1260
actcaccacc aacaccatca ctcgaccggg cactcgatca catcttcata gtctactcaa 1320
atccgcactc aaccagacaa gctgagcaca aggaagagaa gaggagaaga caaagtgttt 1380
ggaagcggcc tggacctcca tcggatcacg aagaagccca tctcggccca ttatcattct 1440
atgggccggg cgattaggtt attggcccgt ctactatcat tttatttcgt tatgtataaa 1500
tagatgtctt agggttctgt accaggacat ctagtcgaca ttgagttttt ttgcttcagt 1560
tttattttct gttttctctg ctgcgccgct tttgtttctg caacctgtaa ttcgagattt 1620
ttccaagtta ctcagattcc gcatttgatt tc                              1652
```

<210> 5
<211> 903
<212> DNA
<213> Arabidopsis thaliana

<400> 5
```
tactctcttc caaatttggt cacacagtgg actgtgtgat ttaagtttgg gggagggctc  60
aggaagtgtg tgttgcattg tataatcttg agtttgcatt catctaaggc atagaaaaac 120
caaaaaaatt gaaaaattcc agaaaatgat ttcacaaaaa tagagtgttc atgtagttgc 180
attgcattta ggatcgagtt tagagtgttt cgtttaggat tgttgcatat gcatagggga 240
taataatgag atagccttgt aagcattttg gttcaccaga taagctcagt gccctcgttg 300
ttagttgttt gatgcgttgt cattgaaatt gaagtaagaa ctgcacgatg cctagattgc 360
tctactcgac cacactgtta ggatctgata tcattcccta tcaatttgaa cttgaatctg 420
atttagaatt atcatgtctt ggcatcgaat ttgaactcat ggatacccta aaatacttgg 480
attttcttac tcattttaac cactcttgtt gatccaagta gctgactctc cttattagag 540
cagttaaccc atacccaaac ctgaactttc tttcaagccc tatatcactt gtgagtgttt 600
gtgaggtctt atttcgattg agcttggtag aaagtgttag gttcgtaacg acagagatag 660
tgtctcatgt agttctagtt tgcgttcttc agactggata ggactaggtg ggcgcttata 720
tcatgggttg ggatgtgttt aaaagaaaag agggaatcta ttgttgatga ggaaagggaa 780
agaattccag gggaagtaag ctaaagaagt tagaaaaaaa atctagtaaa ggttttggga 840
atgttaaaga aaagaatgag gttcttgtta gctaaagaag aagggttaaa agcctttggt 900
ttt                                                              903
```

<210> 6
<211> 614
<212> DNA
<213> Arabidopsis thaliana

<400> 6
```
gagggcaagc aaagactaag tttgggggag ttgataagtg tgtattttgc atgttttgag  60
catccatttg tcatcacttt agcatcatat catcactgtt ttataccatt tcacatcatt 120
tgtcatcact ttgcatgttt aggatagttt tgcatgcatg ttgcatattt gcgttgattt 180
caggtgattt ggagctgttg acgagctatc tggaagaagc agacctgatc atgacaaacc 240
actcgaccca gaggtcgagt aggagcttca agatctcaag agactactcg acaaccaggt 300
cgagtagagc acatcaccac ttcacctcac cactcgaccc cgaggtcgag tgccatcatc 360
tccatcacca gacggtcact cgatcacttc actcgacctt gaggtcgagt gtcttcacct 420
ccatcatcag acaaccactc gacctcctca ctccacctag aggtcgagta tctccatctt 480
accactcgac tgcatactcg atgacaagct tcagagcctt cttaattccg cactcaacca 540
gacactcgag cacaaggaag aaaagaagac tccagctatt cactcgagct ctcactcgac 600
```

```
cacgtgggtc gagt                                                      614


<210> 7
<211> 1956
<212> DNA
<213> Arabidopsis thaliana

<400> 7
ttttggttca ccggattaac tcagtgctct cgttgctagt tgtgtgttgc gtagtgaatg 60
aatttgaaag aaaactgaac catgcctaga ttgctctact cgaccacact gtcatgatct 120
gataccattc cctatcaatt tgaacctgaa tttgatcttt aattatcatg tctgcatcaa 180
atttgaactc atggataccc taaaatactt ggattttctt attcattttg atcactcttg 240
ttaatccaag tagctgactc tccttattag agcagttaac ccgaacccaa acctagactt 300
tttttcaagc cttatatcac tcgtgagggt ttgtgaggtc ttattccgat tcagcttggt 360
agaaagtgtt aggttcgtaa cgacagagat agtgnctcat gtagttctag ttcgcatttt 420
ttggactaga taggactggg tgggcgctta tactttaggt tgggatgngt ttaaaagaaa 480
aaaaaagggg ttgattcatt gatgagaaaa ggtaaaagac tctaggtgaa gtaagataaa 540
gaagcagaaa aggtctagta aaggttttgg gatttgtaaa aaaaagaaag agttcttgtt 600
agctattgaa gatgggcaaa agccctcggt tttaaaatgt taaaaacagg aaccttagtt 660
gttaaagaaa tccaaatccg ctagatgtat caaagtgttg agaaagcttt tcctagagtt 720
aagagaaaag aaaagaatga ttagaaaaag ggcttaaagg attcatgaat gcaaagggta 780
gaggtaagtt cttatactgg gattggagat gggattacca ttagagcttc atctgatata 840
ctctaggtag atgggatctt atctctgcat gcatagtttg ggacttacct ttagcattct 900
actaaagctt aatcattttt tgagagatcc cctgttactg aagcctattc tgtaagggac 960
catctttgtc tcttgacctt ttaccttagc caaatgagtt cattgatgat gcattgcttg 1020
attcacgttc cagaactaat gaatgttaaa gggattggta gatttgaaaa catgtgtagg 1080
tcgagcatat gagtcggatt gattgatagt aaggcatggc taaagttttt cagtagaatt 1140
cgatcatatc gcagcttaga actatcaact tggacattga tttcatttgg tttatctagt 1200
gctttggctc tgagtccccg atttcaaacc tcacctctag cttgttctta attgtttgct 1260
tgagggcaag caaagactaa gtttggggga gttgataagt gtgtattttg catgttttga 1320
gcatccattt gtcatcactt tagcaccata tcatcactat tttataccat ttctcatcat 1380
ttgtcatcac tttgcatgtt taggatagtt ttgcatgcat gtggcatatt tgtgttgttt 1440
tcaagtgatt cggagctgtt gaagaactaa ttggaagaag cggacctgat catgccaaac 1500
cactcgacct caggtcgagt agacgcttca cgacctcaac acaccactcg accacctggt 1560
cgagtgtagg acttcaccac ttcacctcat cactcgaccc cctggccgag taccccacga 1620
gagtcactcg atcacttcac tcgaccccca ggtcgagtgt cttcacctcc accacctgac 1680
catcactcga tcacacgact ctacctggaa gtcgagtatc accatcacca ccactcgact 1740
acatacttga tgtcgagctt cagagtcttc tccattccgc actcaaccag acactcgagc 1800
acaaggaaaa aaagaagatt ctagcttatc actcgacctc tcactcgacc acctgggtcg 1860
agtacagttc ttaatccgtc tcaatactgc gtcgttttga gtattagggt ttcggaatat 1920
ttttgctata agtagcacgt actttacatt ttcgag                             1956


<210> 8
<211> 2105
<212> DNA
<213> Arabidopsis thaliana

<400> 8
ttagcatttt ggttcactag ataaactcag tgccctcgtt gttagttgtc tgatgcatac 60
tcaatgaaat tgaagtaaaa ctgcaccatg cctagattgc tctactcgac cacactgtta 120
ggatctgata ccattcccta tcaatttgaa cttgaatctg atttagaatt atcatgtctt 180
gccatcgaat ttgaactcat ggataccctc aaatacttgg attttcttac tcattttaac 240
cactcttgtt aatccaagta gctgactctc cttattagag cagttaaccc gaatccaaac 300
ctaatctttc tttcgagccc tatatcactt gtgagtgttt gtgaggtctt atttcaattg 360
```

```
agcttggtag aaagtgttag gttcgtaacg acagagatag tgtctcatgt agttctagtt 420
cgcgtttttt ggactggata ggactaggtg ggcgcttata tcatgggttt ggatcagttt 480
aaaaaaaaaa aagggtgaat tcattgttga taaggaaagg gaaagaattc taggggaagt 540
aagctaaaga agttagaaaa aaaaaaatct agtaaaggtt ttgggaatgt taaagaaaag 600
aatgaggttc ttgttagcta aagaagaagg gttaaaagcc ttttgtttta aagattaaaa 660
acaggaacct tagttgttaa agaaatccaa atacgctaga tgtatcagag tgttgagaaa 720
gcttctccta gagttaagag aaaagaaaag aatgatatga aaaagagttt gaaagattca 780
tgagtgcaaa gggtagagtt aagttcttgt attgggactg gagttgggat taccattaga 840
gcttcattgt tatactatgg gtagatggga ttttatctct gtatgcataa cttgggactt 900
acctttagca ttctactaaa gctcaatcat tcttgagaga tcccctgtta cttaagccta 960
ttctgtaagg gaccatcttt gtctcttgcc ttcaccttag ccaaatgagt tcattgatga 1020
tgcattgttt gattcacgtt ccagaactaa tgaatgttaa agggattggt agatttgaaa 1080
gcatgtgtag gtcgagtata agagacggat tgattgataa caaggcatgg ctaacgtttt 1140
cgagtaaaat tcaatcatat cgcatcttag aactaccaac ttggacattg attttatttg 1200
ctctatcaga tgctttggtt ctgagtcccc accttcaaac ctctccttca actatgtctt 1260
cttatttgct tgagggcaag caaagactaa gtttggggga gttgatatgt ctataatttg 1320
catgttttca gtgtccattc atcatcgttt tgagtccagt ttcgtatcat tcatcactgt 1380
tttatatcat ttctcatcat tcttgcatac tttgcatgat taggatagct ttgtatacat 1440
attgcatttc tgagttgttt ttaggtgatt tggagctgtt tgcgagcaaa ttggaagaaa 1500
cgagccagaa caagaagcca tactcgaccc cctggtcgag tgactttgga gccattcttc 1560
ccatctactc gacccggggg tcgagtaacc tcagctcagg ccactcgatg acgccactcg 1620
tccccctg tcgagtatca cttcgccaca ccacctgacc acactcggcc gttcactcta 1680
ccacgttact cgaccccctg gtcgagtatc atcactcacc accaacacca tcactcgacc 1740
gggcactcga tcacatcttc atagtctact caaatccgca ctcaaccaga caagctgagc 1800
acaaggaaga gaagaggaga agacaaagtg tttggaagcg gcctggacct ccatcggatc 1860
acgaagaagc ccatctcggc ccattatcat tctatgggcc gggcgattag gttattggcc 1920
cgtctactat cattttatt cgttatgtat aaatagatgt cttagggttc tgtaccagga 1980
catctagtcg acattgagtt tttttgcttc agttttattt tctgttttct ctgctgcgcc 2040
gcttttgttt ctgcaacctg taattcgaga tttttccaag ttattcagat tccgcatttg 2100
atttc                                                            2105
```

```
<210> 9
<211> 4860
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Virtual TSI

<400> 9
tcgccgttcc agcttcgcat actctctcac cgtctctcgt ttcactcgac cacttcacac 60
ttcgcctcaa catcttcgcc ggagtttctc gccattgtcc gtgcttccgt catctccgtt 120
cactcgacca ccggaccggc ttcaccatct ctcaactatc caccattcac tcgacctcgc 180
cattcactgc gcctccattc gtctctttac tcgactgctc ctcaaaccgc caccgtcttc 240
tctaaattcg ccgtttactc gaccacactg ttacgtctct cattcgtgta cagtcgaccg 300
ctatacccga agccacaata tcactctact cgaccgtttc actcgatcgc gtacttgact 360
ggtttagtgt gtgtgtttat ttgaactaac atattgatat ttggttttga gttacattct 420
ttttcaggga atcaatatga gcaactacag tggcgaatcc tccatggatg cggattacaa 480
cgtcgatgaa gctgaatctt ggtcaactag accagagaga gagcaacagg cttatgagag 540
cttcagagcc gagacccaac gctcagtagc tcgacgcaat gaaaggagag ctgagattgc 600
tagaggaaag agagcgatga ccagcagata tgagttgatc gacgaagata ttgacgtcga 660
gtatgagcct gagtcatggc acagagaaac aaaactgttg aacaagcctg atgaagttac 720
agtggaagag tacatcagac tttcgagct gaacgacttc tggggagcga ggtacccctg 780
ttatgagact ctagcccagc ttaggctact ggaggacgta cagcacttat tcgagaagtg 840
ccatcttgag acgctgatgt cttacccgta cgtcgcttac aagaaggaaa caatagagtt 900
```

```
tctctccact ctgcaagtgg agttgtatca gggacttact gcagatgaac tggagagtga 960
agggttggga ttcttgactt tttcagttaa cgagcagcgt taccagctat ctatcaagag 1020
cttggaagga ttatttggtt ttcccagtgg aaagggaact aaacccaagt tcgaaaggga 1080
agagttgaag gatttgtggt taaccattgg gaacgatttg gcgctcaact ctgcaaggtc 1140
taagagcaac cagattcgaa gccctgtgat ccgctactat cagcgctcag tagcgaatgt 1200
tctgtacccc agggaatcta caggcaccgt gtctaacaca gacatggaga tgattgattc 1260
tgcactcaag ggtattctcc ggagaacaaa ggggaagaag gtcctaaagg gcgaccttaa 1320
tgatacacca ccggtcatgc ttctgttgat ccatatgtgt ggatacagga agtgggcgca 1380
caccaacggg aggaagaagg tgcgaggagc cctttgtgtg ggtggcgttg tgacaccgat 1440
tctgattgca tgtggtgtac ctctcacgtc tccagggttt gatccgagga tgatggattt 1500
agatcacttg cgtcgttgtg agtttctgga gtacgacatg gttggcgatt tctatcgcta 1560
caaattcgag cactccctga cccgaacagc caacattttg cttccctgca tcgaggccac 1620
aaccatactt tagggtgaga acattgactt cagacctgcg cgtgattacc tctactttga 1680
gagcactcca ccgactgatg acaatgtccc tacgacggaa gctacagagg atgattttgc 1740
tgagacggat gaggatagg aggaggagta tgatacgagc atgtatcatt tcagtgagca 1800
cgtacctcca gcgcaggaga gcaagagctt gagtgaagct cacagaaaca acagtaagtt 1860
gcagaggtgg tgcaagaaac aagataggtt acttatcaag tgcttcaagg ccatcacgtt 1920
tctaacggac aagataagtt gcttctcttc taccacagct attccgcagg gagagcgtcc 1980
tcaggacatg ccttcgaaga gatatgacgc gccagggcca agtcatcaca ggcctgagcc 2040
aagtcaccac aggcctgagc ctagtgaccg agtagtacca ccagtccctg caaggcattc 2100
atcattcgag cctcgggagc tcgggagaaa gaagaaggct gcactcgcta ggtctggcag 2160
caggagtaga cgacttctcc agtcccgtag cttacgcgac cgcggtgctg ccgcagcag 2220
aagaagagag gtcgagtatc atcagagcgg tgctggccgc ggcgaaggag cagaggtcga 2280
gtaccccag ggggaagctg agacacaaca gggagattct tcgatggcct gggagcaatc 2340
acaggcagct attgacgacc aactccgctc cttcttccac tgaggtatgc acctcactcc 2400
accattgtaa tataccatct cttgtttttt attttgtttt tgtgatgtgt tttgtcctga 2460
gtactctctt ccaaatttgg tcacacagtg gactgtgtga tttaagtttg ggggagggct 2520
caggaagtgt gtgttgcatt gtataatctt gagtttgcat tcatctaagg catagaaaaa 2580
ccaaaaaaat tgaaaaattc cagaaaatga tttcacaaaa atagagtgtt catgtagttg 2640
cattgcattt aggatcgagt ttagagtgtt tcgtttagga ttgttgcata tgcataggg 2700
ataataatga gatagccttg taagcatttt ggttcaccag ataagctcag tgccctcgtt 2760
gttagttgtt tgatgcgttg tcattgaaat tgaagtaaga actgacgat gcctagattg 2820
ctctactcga ccacactgtt aggatctgat atcattccct atcaatttga acttgaatct 2880
gatttagaat tatcatgtct tggcatcgaa tttgaactca tggataccct aaaatacttg 2940
gattttctta ctcattttaa ccactcttgt tgatccaagt agctgactct ccttattaga 3000
gcagttaacc catacccaaa cctgaacttt ctttcaagcc ctatatcact tgtgagtgtt 3060
tgtgaggtct tatttcgatt gagcttggta gaaagtgtta ggttcgtaac gacagagata 3120
gtgtctcatg tagttctagt ttgcgttctt cagactggat aggactaggt gggcgcttat 3180
atcatgggtt gggatgtgtt taaaagaaaa gagggaatct attgttgatg aggaaaggga 3240
aagaattcca ggggaagtaa gctaaagaag ttagaaaaaa aatctagtaa aggttttggg 3300
aatgttaaag aaaagaatga ggttcttgtt agctaaagaa gaagggttaa aagcctttgg 3360
ttttaaagat taaaaaaaa acaggaacct tagttgttaa agaaatccaa acccgctaga 3420
tgtatcaaga gcgttgagaa agcttctcct agagttaaga gaaaagaaaa gaatgatatg 3480
aaaaagagtt tgaaagattc atgagtgcaa agggtagagt taagttggga caggagttgg 3540
ttttaccatt agaacttcat tgttatactc tgggtagatg ggatcttatc tctgtatgca 3600
taatttggga cttacctta gcattctact aaagctcaat cattcttgag ggatcccctg 3660
ttacttaagc ctattctata agggaccatc tttgtctctt gaccttcacc ttggccgaat 3720
gagttcattg atgatgcatt gcttgattcg cgttccagaa ctaatgaatg ttaaagggat 3780
tggtagattt gaaagcatgt gtaggtcgag tataagagac ggattgattg aaaacaaggc 3840
atggctaacg tttttgagta gaattcaatc atatcgcatc ttagaactac caacttggac 3900
attgatttta tttgctctat catatgcttt ggttttgagt ccccgccttc actcctctcc 3960
ttcaactatg tcttcttatt tgcttgaggg caagcaaaga ctaagtttga gggagttgat 4020
atgtctataa tttgcatgtt ttcagtgtcc attcatcatc gtttttgagtc cagtttcgta 4080
tcattcatca ctgtttata tcatttctca tcattcttgc atactttgca tgattaggat 4140
aactttgcat acatattgca tttctgagtt gtttttcaggt gatttggagc tgtttgcaag 4200
caaattggaa gaaatgagcc agaaccagaa gacatactcg acccctaggt cgagtgactt 4260
```

```
tggggccatt cttcccacat actcggcccc caggtcgagt gactttggag ccattcttcc 4320
catccactcg accaccgggt cgagtaacct tagctcaggc cactcgatga cactactcga 4380
ccccaggtcg agtatcactt cgccacacca cctgacaaca ctcgaccaat cactctacca 4440
agttactcga cccctggtc gagtatcatc actcaccacc atcagcatca ctcgaccgga 4500
cactcgatca cgtcttcaca gtctactcaa atccgcagtc aaccagacaa gctgagcaca 4560
aggaagagaa gaggagaaga caaagtgctt ggaagcggcc tggacctcca tcggatcacg 4620
aagcccatct cggcccatta tctctctatg ggccgagcga ttaggttatt ggcccgtcta 4680
ctatcatttt atttcgtttt gtataaatag atgtcttagg gttttgtcct gagacatcta 4740
gtcgacattg agttttttt gcttcagttt tattttctgt tctactctgc tgcgccgctt 4800
ttgcttctgc aacctgtaat tcgagatttt tccaagttat tcagattccg catttgattt 4860
```

<210> 10
<211> 648
<212> PRT
<213> Arabidopsis thaliana

<400> 10

```
Met Ser Asn Tyr Ser Gly Glu Ser Ser Met Asp Ala Asp Tyr Asn Val
  1               5                  10                  15

Asp Glu Ala Glu Ser Trp Ser Thr Arg Pro Glu Arg Glu Gln Gln Ala
             20                  25                  30

Tyr Glu Ser Phe Arg Ala Glu Thr Gln Arg Ser Val Ala Arg Arg Asn
         35                  40                  45

Glu Arg Arg Ala Glu Ile Ala Arg Gly Lys Arg Ala Met Thr Ser Arg
     50                  55                  60

Tyr Glu Leu Ile Asp Glu Asp Ile Asp Val Glu Tyr Glu Pro Glu Ser
 65                  70                  75                  80

Trp His Arg Glu Thr Lys Leu Leu Asn Lys Pro Asp Glu Val Thr Val
                 85                  90                  95

Glu Glu Tyr Ile Arg Leu Phe Glu Leu Asn Asp Phe Trp Gly Ala Arg
                100                 105                 110

Tyr Pro Cys Tyr Glu Thr Leu Ala Gln Leu Arg Leu Leu Glu Asp Val
            115                 120                 125

Gln His Leu Phe Glu Lys Cys His Leu Glu Thr Leu Met Ser Tyr Pro
    130                 135                 140

Tyr Val Ala Tyr Lys Lys Glu Thr Ile Glu Phe Leu Ser Thr Leu Gln
145                 150                 155                 160

Val Glu Leu Tyr Gln Gly Leu Thr Ala Asp Glu Leu Glu Ser Glu Gly
                165                 170                 175

Leu Gly Phe Leu Thr Phe Ser Val Asn Glu Gln Arg Tyr Gln Leu Ser
            180                 185                 190

Ile Lys Ser Leu Glu Gly Leu Phe Gly Phe Pro Ser Gly Lys Gly Thr
        195                 200                 205
```

Lys Pro Lys Phe Glu Arg Glu Glu Leu Lys Asp Leu Trp Leu Thr Ile
    210   215   220

Gly Asn Asp Leu Ala Leu Asn Ser Ala Arg Ser Lys Ser Asn Gln Ile
225   230   235   240

Arg Ser Pro Val Ile Arg Tyr Tyr Gln Arg Ser Val Ala Asn Val Leu
            245   250   255

Tyr Pro Arg Glu Ser Thr Gly Thr Val Ser Asn Thr Asp Met Glu Met
        260   265   270

Ile Asp Ser Ala Leu Lys Gly Ile Leu Arg Arg Thr Lys Gly Lys Lys
        275   280   285

Val Leu Lys Gly Asp Leu Asn Asp Thr Pro Pro Val Met Leu Leu Leu
    290   295   300

Ile His Met Cys Gly Tyr Arg Lys Trp Ala His Thr Asn Gly Arg Lys
305   310   315   320

Lys Val Arg Gly Ala Leu Cys Val Gly Gly Val Val Thr Pro Ile Leu
            325   330   335

Ile Ala Cys Gly Val Pro Leu Thr Ser Pro Gly Phe Asp Pro Arg Met
        340   345   350

Met Asp Leu Asp His Leu Arg Arg Cys Glu Phe Leu Glu Tyr Asp Met
        355   360   365

Val Gly Asp Phe Tyr Arg Tyr Lys Phe Glu His Ser Leu Thr Arg Thr
    370   375   380

Ala Asn Ile Leu Leu Pro Cys Ile Glu Ala Thr Thr Ile Leu Xaa Gly
385   390   395   400

Glu Asn Ile Asp Phe Arg Pro Ala Arg Asp Tyr Leu Tyr Phe Glu Ser
            405   410   415

Thr Pro Pro Thr Asp Asp Asn Val Pro Thr Thr Glu Ala Thr Glu Asp
            420   425   430

Asp Phe Ala Glu Thr Asp Glu Asp Arg Glu Glu Glu Tyr Asp Thr Ser
        435   440   445

Met Tyr His Phe Ser Glu His Val Pro Pro Ala Gln Glu Ser Lys Ser
    450   455   460

Leu Ser Glu Ala His Arg Asn Asn Ser Lys Leu Gln Arg Trp Cys Lys
465   470   475   480

Lys Gln Asp Arg Leu Leu Ile Lys Cys Phe Lys Ala Ile Thr Phe Leu
            485   490   495

Thr Asp Lys Ile Ser Cys Phe Ser Ser Thr Thr Ala Ile Pro Gln Gly

```
              500                    505                    510
```

Glu Arg Pro Gln Asp Met Pro Ser Lys Arg Tyr Asp Ala Pro Gly Pro
        515   520   525

Ser His His Arg Pro Glu Pro Ser His His Arg Pro Glu Pro Ser Asp
    530   535   540

Arg Val Val Pro Pro Val Pro Ala Arg His Ser Ser Phe Glu Pro Arg
545   550   555   560

Glu Leu Gly Arg Lys Lys Lys Ala Ala Leu Ala Arg Ser Gly Ser Arg
                565   570   575

Ser Arg Arg Leu Leu Gln Ser Arg Ser Leu Arg Asp Arg Gly Ala Gly
            580   585   590

Arg Ser Arg Arg Arg Glu Val Glu Tyr His Gln Ser Gly Ala Gly Arg
        595   600   605

Gly Glu Gly Ala Glu Val Glu Tyr Pro Gln Gly Glu Ala Glu Thr Gln
    610   615   620

Gln Gly Asp Ser Ser Met Ala Trp Glu Gln Ser Gln Ala Ala Ile Asp
625   630   635   640

Asp Gln Leu Arg Ser Phe Phe His
                645

<210> 11
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 11
tggttcacca gataagctca gtgccctc   28

<210> 12
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 12
cttcagactg gataggacta ggtgggcg   28

                                    - 10 -
```

<210> 13
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 13
cgcccaccta gtcctatcca gtctgaag                                    28


<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 14
cgcatcaaac aactaacaac gagggcac                                    28


<210> 15
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 15
cgataacatc gaccgtattg ctcgcc                                      26


<210> 16
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 16
aactagctcc catccgtctt cgacatcc                                    28


<210> 17
<211> 24
<212> DNA

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
        Oligonucleotide sequence

<400> 17
tgcatcacac cggattggat tgac                                                24


<210> 18
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
        Oligonucleotide sequence

<400> 18
tgttcccctg aaccatagca atgagacc                                            28


<210> 19
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
        Oligonucleotide sequence

<400> 19
caaacagaca gagtgtggcc caccacc                                             27


<210> 20
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
        Oligonucleotide sequence

<400> 20
caaacagaca gagtgtggcc caccacc                                             27


<210> 21
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:

Oligonucleotide sequence

<400> 21
tgcaaaccca caggaccaag tctaccc                                          27


<210> 22
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 22
acagatggtg atagcgtgag cggtggc                                         27


<210> 23
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      oligonucleotide sequence

<400> 23
tcaacctttt gccccaacaa ccactc                                          26


<210> 24
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      oligonucleotide sequence

<400> 24
tctccatcca cgctttcctg aatgtcc                                         27


<210> 25
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      oligonucleotide sequence

<400> 25
ggagaaggaa gctgaaaatc atattgtgg                                       29

<210> 26
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      oligonucleotide sequence

<400> 26
atgatgatcc taagtctacc cttttgcac                                          29

<210> 27
<211> 3122
<212> DNA
<213> Arabidopsis thaliana

<400> 27
ttcatatatt cgacctcttc ttctcattct tgcatccaaa agacacaaca agccgccatc 60
gctttccctc acaactctca ctcgaccacc gcccgctctc tcacttactc ggcttcatcg 120
ctctcatcgc catctctcaa catactcgac ctcgcgatat cactcgagct cgccgcttct 180
caccgcctct ccatcgtcac cgcctgctcc ctctctccaa ggaaacaact cgagctctcc 240
atttcactca ctcgacctct accaccaagc cggcttcacc acttctagct cttaaccact 300
cgaccacctt caccatcaac caatcaaatc gttttctcct ccattaaagc ttgacatact 360
cgaccgctga acacttatca ccttcaagct cctcatctct tcatcgtttc caacaccgct 420
gctctcatcc cccacgaaag cttgtcatca cctctcactc atcaccagtt cactcgattc 480
agcaaccaaa ctcgacctcg tctctcttgc cactcatagt cactcgatct ctcctcacca 540
tcttcatcat ctcccttact cgaccaccgt gcgtctcgct ccaccattgc catttaaaag 600
ctcactcgat tgtcaaagag aagaagagtg aagctcaacc accgccactc gaccgcgttt 660
ccctctacac attcaacact cgaccacggt gctaccatct ccacacccgc tcttgttcac 720
catacactcg accaacaact ctcaaagtaa aaaaaaaaag aaaaaaaaag tcaaaaccga 780
cagtttcact caaccggttt actcgaccgg tacgctggtt tagattgtgt ttttggtttt 840
gctattacta acatattaac gtttatcttt gagtttcgtc tgtttttagg tttcatcatg 900
agtaactaca gtggaaaatc ctctatggac cctgattata atgtggatga agctaagtcc 960
tggtccacta gaccggagtg agagcaacat gtttacgaga gctataggga tgaatttgaa 1020
cgctctgcag ctcgacgtaa tcaaagaaga gctgaaatcg ctagaggaaa gagggcgatg 1080
tcgagtagat atgagctgat tgatgaggat atcaaaactg agtatgagcc agagtcatgg 1140
cgcaaggaga cgaagctact gaacaaatcc gacgaggtta cagtggagga gtatatcaga 1200
ttctttgaga tgaatgactt ctggggaacg aggtatccct gatatgagac tttagcccag 1260
ttggggttac tggaggacgt gcagcatctg ttcgagaagt gtcatctgat aaggaggaga 1320
caatcgagtt tctttccaca ctgcaagtgg aaatgtatga gggactcaca gactttgagc 1380
tggataccat ggggttaggc ttcttgacgt tcttagtgga tgaacagcgg taccagattt 1440
agatcaagaa attggaagaa ctgtttggtt tccctagtgg aaagggaacc aaccccaggt 1500
ttgacaggga gagcttaag gatttgtggg ctactattgg gaacaatcta ccgctaaact 1560
cgacgcggtc aagagcaac caaatccgga gtcctgtgat tcgctacttt cagcgctcgg 1620
ttgccaatgt tttttactcc agggagtcta caggcaccgt gtctaacaca gacatgaaga 1680
tgatagattc agcgcttata gggattctcc gccttacaaa aggaaagaat gtcctgagag 1740
gagatcttaa cgactcacca ccagtaatgc ctctgttgat ccatctgtgt gggtacatga 1800
agtgggcgct gacaaacggc aagaagaagg taagaggagc actatgcgtg ggtggcgttg 1860
tgacgccaat tctgaaagtt tgtggagttc cgctcaagga agtagggtta gcaccgagaa 1920
tgatggactt ggatcacttg cgccgatgtg agttctctga gtttgacatg gttggcgact 1980
ttcaccgcta caggttcgag cattcatcga ttagaatcgc caacattctt ttcccctgca 2040
tttacgctac taggattctc gagggcagga acattgactt caagcctgcg cttgaagatc 2100

```
tttatttcga gggcagtccg ccaactgagg agattagtca caccgaagga gctacaatag 2160
aagatgttga tgagacatat gatatagatg aggcggagtt tgacacgagc atgtatcatt 2220
tcagtgagca tatacctcca gcgaggaaaa gcaagagttt gagcgaagct cacaggaaca 2280
acagcaagct gcagaagtgg tgcaagaaac aggataagtt actcgccaag tgcctcaggg 2340
ctatcaagtt tctgaaggac aagatcagct gctcctcttc cactacaact attccgcaat 2400
gacagctccc tcaggacatg ccttcgagga gatatgacgc gcccgagcct agagagcaga 2460
agattctgca tgtccctgcg aggcattcat cattcgagcc tcgtgaatct aggaagaata 2520
ggagaacgac actcactcga tctagcagca ggagcagacg acttctgcag tctcgtagtt 2580
tacgcgaccg cggtgctggc cgcaatagaa gaagagaggt cgagtatcct cagagcggtg 2640
ctggccgcca cagagctgat gagatcgagt acccacatgc tggagctgat acggaacatg 2700
gcggttcgtc tatggcttgg gagcaatcac aggcagccat tgactaccaa cttcgttcat 2760
tattcgactg aggtaagcgc ctcacttcac cattatatta tatcatctct tgtgatttgt 2820
tctttatttt gtttcagtga ttggatttgt cctgagtact ctcttccaag tttattcaca 2880
cagtggactg tgtgatttaa gtttggggga gggctcagga agtatgttgc attgtatata 2940
tttttaagtc tgcattcatc taaggcatag aaaaaccaaa aaaaaattaa aaatttcaga 3000
aaatgatttc acaaaaaaag agtgttcatg tagttgcatt acatttagga tcaagtctag 3060
agtgtttcat ttaggattgt tgcatatgca tagggataa tgatgagata gccttgtaag 3120
ca                                                                  3122
```

31147
SEQUENCE LISTING

<110> Syngenta Participations AG

<120> Transcriptionally Silenced Plant Genes

<130> S-31147A

<140>
<141>

<150> GB 9921964.4
<151> 1999-09-16

<160> 26

<170> PatentIn Ver. 2.1

<210> 1
<211> 2512
<212> DNA
<213> Arabidopsis thaliana

<400> 1

```
tcgccgttcc agcttcgcat actctctcac cgtctctcgt ttcactcgac cacttcacac 60
ttcgcctcaa catcttcgcc ggagtttctc gccattgtcc gtgcttccgt catctccgtt 120
cactcgacca ccggaccggc ttcaccatct ctcaactatc caccattcac tcgacctcgc 180
cattcactgc gcctccattc gtctctttac tcgactgctc ctcaaaccgc caccgtcttc 240
tctaaattcg ccgtttactc gaccacactg ttacgtctct cattcgtgta cagtcgaccg 300
ctatacccga agccacaata tcactctact cgaccgtttc actcgatcgc gtacttgact 360
ggtttagtgt gtgtgtttat ttgaactaac atattgatat ttggttttga gttacattct 420
ttttcaggga atcaatatga gcaactacag tggcgaatcc tccatggatg cggattacaa 480
cgtcgatgaa gctgaatctt ggtcaactag accagagaga gagcaacagg cttatgagag 540
cttcagagcc gagacccaac gctcagtagc tcgacgcaat gaaaggagag ctgagattgc 600
tagaggaaag agagcgatga ccagcagata tgagttgatc gacgaagata ttgacgtcga 660
gtatgagcct gagtcatggc acagagaaac aaaactgttg aacaagcctg atgaagttac 720
agtggaagag tacatcagac ttttcgagct gaacgacttc tggggagcga ggtacccctg 780
ttatgagact ctagcccagc ttaggctact ggaggacgta cagcacttat tcgagaagtg 840
ccatcttgag acgctgatgt cttacccgta cgtcgcttac aagaaggaaa caatagagtt 900
tctctccact ctgcaagtgg agttgtatca gggacttact gcagatgaac tggagagtga 960
agggttggga ttcttgactt tttcagttaa cgagcagcgt taccagctat ctatcaagag 1020
cttggaagga ttatttggtt ttcccagtgg aaagggaact aaacccaagt tcgaaaggga 1080
agagttgaag gatttgtggt taaccattgg gaacgatttg gcgctcaact ctgcaaggtc 1140
taagagcaac cagattcgaa gccctgtgat ccgctactat cagcgctcag tagcgaatgt 1200
tctgtacccc agggaatcta caggcaccgt gtctaacaca gacatggaga tgattgattc 1260
tgcactcaag ggtattctcc ggagaacaaa ggggaagaag gtcctaaagg gcgaccttaa 1320
tgatacacca ccggtcatgc ttctgttgat ccatatgtgt ggatacagga agtgggcgca 1380
caccaacggg aggaagaagg tgcgaggagc cctttgtgtg ggtggcgttg tgacaccgat 1440
tctgattgca tgtggtgtac ctctcacgtc tccaggtgtt gatccgagga tgatggattt 1500
agatcacttg cgtcgttgtg agtttctgga gtacgacatg gttggcgatt tctatcgcta 1560
caaattcgag cactccctga cccgaacagc caacattttg cttccctgca tcgaggccac 1620
aaccatactt tagggtgaga acattgactt cagacctgcg cgtgattacc tctactttga 1680
gagcactcca ccgactgatg acaatgtccc tacgacggaa gctacagagg atgattttgc 1740
tgagacggat gaggataggg aggaggagta tgatacgagc atgtatcatt tcagtgagca 1800
cgtacctcca gcgcaggaga gcaagagctt gagtgaagct cacagaaaca acagtaagtt 1860
gcagaggtgg tgcaagaaac aagataggtt acttatcaag tgcttcaagg ccatcacgtt 1920
tctaacggac aagataagtt gcttctcttc taccacagct attccgcagg gagagcgtcc 1980
tcaggacatg ccttcgaaga gatatgacgc gccagggcca agtcatcaca ggcctgagcc 2040
aagtcaccac aggcctgagc ctagtgaccg agtagtacca ccagtccctg caaggcattc 2100
atcattcgag cctcgggagc tcgggagaaa gaagaaggct gcactcgcta ggtctggcag 2160
caggagtaga cgacttctcc agtcccgtag cttacgcgac cgcggtgctg gccgcagcag 2220
aagaagagag gtcgagtatc atcagagcgg tgctggccgc ggcgaaggag cagaggtcga 2280
```

Page 1

31147
```
gtacccccag ggggaagctg agacacaaca gggagattct tcgatggcct gggagcaatc 2340
acaggcagct attgacgacc aactccgctc cttcttccac tgaggtatgc acctcactcc 2400
accattgtaa tataccatct cttgtttttt attttgtttt tgtgatgtgt tttgtcctga 2460
gtactctctt ccaaatttgg tcacacagtg gactgtgtga tttaagtttg gg          2512
```

```
<210> 2
<211> 1997
<212> DNA
<213> Arabidopsis thaliana
```

```
<400> 2
ccgggcaggt caacaggctt atgagagctt tagagctgag acccaacgct cagtagctcg 60
acgcaatgaa aggagagctg agattgctag aggaaagaga gcaatgacca gcagatatga 120
gttgatcgac gaagatattg acgtcgagta tgaacctgaa tcatggcaca gagaaacgaa 180
gctgttgaac aagcccgatg aagttacagt agaggagtac atcagacttt tcgagctgaa 240
cgacttctag ggaacgaggt acccctgtta tgagacttta gcccagctgg ggctactgga 300
ggacgtacat cacttattcg agaagtgcca tctggagacg ctgatgtctt acccgtacgt 360
cgcttacaag aaggaaacaa tagagtttct ctccactctg caagtggaga tgtatcaggg 420
acttactgca gatgagctgg agagtgaagg gttggggttc ttgacttttt cagttaacga 480
gcagcgttac cagctatcta tcaagagctt ggaaggatta tttggttttc caagtggaaa 540
gggaactaaa cccaagttcg agagggaaga gttgaaagat ttgtggttaa ccattgggaa 600
cgatttggca ctcaactctg caaggtctaa gagcaaccag attcgaagcc ctgtgatccg 660
ctactatcag cgctcagtag cgaatgttct gtaccccagg gaatctacag gcatcgtgtc 720
taacacagac atggagatga ttgatgttgc actcaagggc attctccgga gaacaaaggg 780
gaagaaggtc ctaaagggcg accttaatga tacaccaccg gttatgcttc tgttgatcca 840
cctgtgtgga tacaggaagt gggcgcacac caacgagaag aagaaggtgc gaggagccct 900
ttgtgtaggt ggcgttgtga caccgattct gattgcatgt ggtgtacctc tcacgtctcc 960
agggtttgat ccgaggatga tggatttaga tcacttgcgt cgttgtgagt ttctagagta 1020
cgacatggtt ggcgatttct atcgctacaa attcgagcac tccctgaccc gaacagccaa 1080
cattttgctt ccctgcatcg aggccacaac catacttcag ggtgagaaca ttgacttcag 1140
acctgcgcgt gattacctct actttgagag cgctccaccg actgatgaca atgtccctac 1200
gacggaagtt acagaggatg atattgctga gacggatgag gataggggag aggagtatga 1260
tacgagcatg tatcatttca gtgagcacgt acctccagcg cgggagagca agagcttgag 1320
tgaagctcac agaaacaaca gtaagttgca gaggtggtgc aagaaacaag ataggctact 1380
tatcaagtgc ttcaaagcca tcacgtttct gacggacaag ataagctgct tctcttctac 1440
cacagctatt ccgcagggag agcatcctca ggacatgcct tcaaggagat atgacgcgcc 1500
agcgccaagt catcacaggc ctgagccaag tcaccacagg cctgagccta gtgaccgagt 1560
agtcccacca gtccctgcaa ggcattcatc attcgagcct cgggagctcg ggagaaagaa 1620
gaaggctgca ctcgctcggt ctggcagcag gagtacacga cttctacagt cccgtagctt 1680
acgcgaccgt ggtgctggcc gcagcagaag aagagaggtc gagtatcatc agagcggtgc 1740
tggccgcgac gaaggagcag aggtcgagta cccacggg aagctgagac acaacaggga 1800
gattcttcga tggcctggga gcaatcacag gcagctattg acgaccaact ccgctccttc 1860
ttccactgag gtatgcacct cactccacca ttgtaatata ccatctcttg tttttatttt 1920
gttttgtga tgtgttttgt cctgagtact ctcttccaaa tttggtcaca cagtggactg 1980
tgtgatttaa gtttggg                                                 1997
```

```
<210> 3
<211> 1682
<212> DNA
<213> Arabidopsis thaliana
```

```
<400> 3
cttatattat gggttgggat gtgtttaaag aaaagggggga attcattgtt gataaggaaa 60
gggaaagaat tctaggggaa gtaagctaaa gaagttagaa aaaatctagt aaaggttttg 120
ggaatgttaa agaaaagaat gaggttcttg ttagctaaag aataagggtt aaaagccttt 180
ggttttaaag attaaaaaaa aaacaggaac cttagttgtt aaagaaatcc aaacccgcta 240
gatgtatcaa gagcgttgag aaagcttctc ctagagttaa gagaaaagaa aagaatgata 300
tgaaaaagag tttgaaagat tcatgagtgc aaagggtaga gttaagttgg gacaggagtt 360
ggttttacca ttagaacttc attgttatac tctgggtaga tgggatctta tctctgtatg 420
cataatttgg gacttacctt tagcattcta ctaaagctca atcattcttg agggatcccc 480
tgttacttaa gcctattcta taagggacca tctttgtctc ttgaccttca ccttggccga 540
```

Page 2

31147
```
atgagttcat tgatgatgca ttgcttgatt cgcgttccag aactaatgaa tgttaaaggg 600
attggtagat ttgaaagcat gtgtaggtcg agtataagag acggattgat tgaaaacaag 660
gcatggctaa cgttttttgag tagaattcaa tcatatcgca tcttagaact accaacttgg 720
acattgattt tatttgctct atcatatgct ttggtttttga gtccccgcct tcactcctct 780
ccttcaacta tgtcttctta tttgcttgag ggcaagcaaa gactaagttt gagggagttg 840
atatgtctat aatttgcatg ttttcagtgt ccattcatca tcgttttgag tccagtttcg 900
tatcattcat cactgtttta tatcatttct catcattctt gcatactttg catgattagg 960
ataactttgc atacatattg catttctgag ttgtttcag gtgatttgga gctgtttgca 1020
agcaaattgg aagaaatgag ccagaaccag aagacatact cgacccctag gtcgagtgac 1080
tttggggcca ttcttcccac atactcggcc cccaggtcga gtgactttgg agccattctt 1140
cccatccact cgaccaccgg gtcgagtaac cttagctcag gccactcgat gacactactc 1200
gaccccaggt cgagtatcac ttcgccacac cacctgacaa cactcgacca atcactctac 1260
caagttactc gaccccctgg tcgagtatca tcactcacca ccatcagcat cactcgaccg 1320
gacactcgat cacgtcttca cagtctactc aaatccgcag tcaaccagac aagctgagca 1380
caaggaagag aagaggagaa gacaaagtgc ttggaagcgg cctggacctc catcggatca 1440
cgaagcccat ctcggcccat tatctctcta tgggccgagc gattaggtta ttggcccgtc 1500
tactatcatt ttatttcgtt ttgtataaat agatgtctta gggttttgtc ctgagacatc 1560
tagtcgacat tgagtttttt ttgcttcagt tttattttct gttctactct gctgcgccgc 1620
ttttgcttct gcaacctgta attcgagatt tttccaagtt attcagattc cgcatttgat 1680
tt                                                                 1682
```

<210> 4
<211> 1652
<212> DNA
<213> Arabidopsis thaliana

<400> 4
```
cttatatcat gggtttggat cagtttaaaa aaaaaaaagg gtgaattcat tgttgataag 60
gaaagggaaa gaattctagg ggaagtaagc taaagaagtt agaaaaaaaa aaatctagta 120
aaggttttgg gaatgttaaa gaaaagaatg aggttcttgt tagctaaaga agaagggtta 180
aaagcctttt gtttttaaaga ttaaaaacag gaaccttagt tgttaaagaa atccaaatac 240
gctagatgta tcagagtgtt gagaaagctt ctcctagagt taagagaaaa gaaaagaatg 300
atatgaaaaa gagtttgaaa gattcatgag tgcaaagggt agagttaagt tcttgtattg 360
ggactggagt tgggattacc attagagctt cattgttata ctatgggtag atgggatttt 420
atctctgtat gcataacttg ggacttacct ttagcattct actaaagctc aatcattctt 480
gagagatccc ctgttactta agcctattct gtaagggacc atctttgtct cttgaccttc 540
accttagcca aatgagttca ttgatgatgc attgtttgat tcacgttcca gaactaatga 600
atgttaaagg gattggtaga tttgaaagca tgtgtaggtc gagtataaga gacggattga 660
ttgataacaa ggcatggcta acgttttcga gtaaaattca atcatatcgc atcttagaac 720
taccaacttg gacattgatt ttatttgctc tatcagatgc tttggttctg agtccccacc 780
ttcaaacctc tccttcaact atgtcttctt atttgcttga gggcaagcaa agactaagtt 840
tgggggagtt gatatgtcta aatttgcatg ttttcagtgt ccattcatca tcgtttttga 900
gtccagtttc gtatcattca tcactgtttt atatcatttc tcatcattct tgcatacttt 960
gcatgattag gatagctttg tatacatatt gcatttctga gttgtttta ggtgatttgg 1020
agctgtttgc gagcaaattg gaagaaacga gccagaacaa gaagccatac tcgacccct 1080
ggtcgagtga ctttggagcc attcttccca tctactcgac ccggggggtcg agtaacctca 1140
gctcaggcca ctcgatgacg ccactcgtcc ccctggtcg agtatcactt cgccacacca 1200
cctgaccaca ctcggccgtt cactctacca cgttactcga cccctggtc gagtatcatc 1260
actcaccacc aacaccatca ctcgaccggg cactcgatca catcttcata gtctactcaa 1320
atccgcactc aaccagacaa gctgagcaca aggaagagaa gaggagaaga caaagtgttt 1380
ggaagcggcc tggacctcca tcggatcacg aagaagccca tctcggccca ttatcattct 1440
atgggccggg cgattaggtt attggcccgt ctactatcat tttatttcgt tatgtataaa 1500
tagatgtctt agggttctgt accaggacat ctagtcgaca ttgagttttt ttgcttcagt 1560
tttattttct gttttctctg ctgcgccgct tttgtttctg caacctgtaa ttcgagattt 1620
ttccaagtta ctcagattcc gcatttgatt tc                                1652
```

<210> 5
<211> 903
<212> DNA
<213> Arabidopsis thaliana

Page 3

31147

<400> 5
```
tactctcttc caaatttggt cacacagtgg actgtgtgat ttaagtttgg gggagggctc 60
aggaagtgtg tgttgcattg tataatcttg agtttgcatt catctaaggc atagaaaaac 120
caaaaaaatt gaaaaattcc agaaaatgat ttcacaaaaa tagagtgttc atgtagttgc 180
attgcattta ggatcgagtt tagagtgttt cgtttaggat tgttgcatat gcatagggga 240
taataatgag atagccttgt aagcattttg gttcaccaga taagctcagt gccctcgttg 300
ttagttgttt gatgcgttgt cattgaaatt gaagtaagaa ctgcacgatg cctagattgc 360
tctactcgac cacactgtta ggatctgata tcattccta tcaatttgaa cttgaatctg 420
atttagaatt atcatgtctt ggcatcgaat ttgaactcat ggataccta aaatacttgg 480
attttcttac tcattttaac cactcttgtt gatccaagta gctgactctc cttattagag 540
cagttaaccc atacccaaac ctgaactttc tttcaagccc tatatcactt gtgagtgttt 600
gtgaggtctt atttcgattg agcttggtag aaagtgttag gttcgtaacg acagagatag 660
tgtctcatgt agttctagtt tgcgttcttc agactggata ggactaggtg ggcgcttata 720
tcatgggttg ggatgtgttt aaaagaaaag agggaatcta ttgttgatga ggaaagggaa 780
agaattccag gggaagtaag ctaaagaagt tagaaaaaaa atctagtaaa ggttttggga 840
atgttaaaga aaagaatgag gttcttgtta gctaaagaag aagggttaaa agcctttggt 900
ttt 903
```

<210> 6
<211> 614
<212> DNA
<213> Arabidopsis thaliana

<400> 6
```
gagggcaagc aaagactaag tttgggggag ttgataagtg tgtattttgc atgttttgag 60
catccatttg tcatcacttt agcatcatat catcactgtt ttataccatt tcacatcatt 120
tgtcatcact ttgcatgttt aggatagttt tgcatgcatg ttgcatattt gcgttgattt 180
caggtgattt ggagctgttg acgagctatc tggaagaagc agacctgatc atgacaaacc 240
actcgaccca gaggtcgagt aggagcttca agatctcaag agactactcg acaaccaggt 300
cgagtagagc acatcaccac ttcacctcac cactcgaccc cgaggtcgag tgccatcatc 360
tccatcacca gacggtcact cgatcacttc actcgacctt gaggtcgagt gtcttcacct 420
ccatcatcag acaaccactc gacctcctca ctccacctag aggtcgagta tctccatctt 480
accactcgac tgcatactcg atgacaagct tcagagcctt cttaattccg cactcaacca 540
gacactcgag cacaaggaag aaaagaagac tccagctatt cactcgagct ctcactcgac 600
cacgtgggtc gagt 614
```

<210> 7
<211> 1956
<212> DNA
<213> Arabidopsis thaliana

<400> 7
```
ttttggttca ccggattaac tcagtgctct cgttgctagt tgtgtgttgc gtagtgaatg 60
aatttgaaag aaaactgaac catgcctaga ttgctctact cgaccacact gtcatgatct 120
gataccattc cctatcaatt tgaacctgaa tttgatcttt aattatcatg tctgcatcaa 180
atttgaactc atggataccc taaaatactt ggattttctt attcattttg atcactcttg 240
ttaatccaag tagctgactc tccttattag agcagttaac ccgaacccaa acctagactt 300
tttttcaagc cttatatcac tcgtgagggt ttgtgaggtc ttattccgat tcagcttggt 360
agaaagtgtt aggttcgtaa cgacagagat agtgnctcat gtagttctag ttcgcatttt 420
ttggactaga taggactggg tgggcgctta tactttaggt tgggatgngt ttaaaagaaa 480
aaaaaagggg ttgattcatt gatgagaaaa ggtaaaagac tctaggtgaa gtaagataaa 540
gaagcagaaa aggtctagta aaggttttgg gatttgtaaa aaaaagaaag agttcttgtt 600
agctattgaa gatgggcaaa agccctcggt tttaaaatgt taaaaacagg aaccttagtt 660
gttaaagaaa tccaaatccg ctagatgtat caaagtgttg agaaagcttt tcctagagtt 720
aagagaaaag aaaagaatga ttagaaaaag ggcttaaagg attcatgaat gcaaagggta 780
gaggtaagtt cttatactgg gattggagat gggattacca ttagagcttc atctgatata 840
ctctaggtag atgggatctt atctctgcat gcatagtttg ggacttacct ttagcattct 900
actaaagctt aatcattttt tgagagatcc cctgttactg aagcctattc tgtaagggac 960
catctttgtc tcttgacctt ttaccttagc caaatgagtt cattgatgat gcattgcttg 1020
attcacgttc cagaactaat gaatgttaaa gggattggta gatttgaaaa catgtgtagg 1080
tcgagcatat gagtcggatt gattgatagt aaggcatggc taaagttttt cagtagaatt 1140
```
Page 4

```
cgatcatatc gcagcttaga actatcaact tggacattga tttcatttgg tttatctagt 1200
gctttggctc tgagtccccg atttcaaacc tcacctctag cttgttctta attgtttgct 1260
tgagggcaag caaagactaa gtttgggggga gttgataagt gtgtattttg catgtttttga 1320
gcatccattt gtcatcactt tagcaccata tcatcactat tttataccat ttctcatcat 1380
ttgtcatcac tttgcatgtt taggatagtt ttgcatgcat gtggcatatt tgtgttgttt 1440
tcaagtgatt cggagctgtt gaagaactaa ttggaagaag cggacctgat catgccaaac 1500
cactcgacct caggtcgagt agacgcttca cgacctcaac acaccactcg accacctggt 1560
cgagtgtagg acttcaccac ttcacctcat cactcgaccc cctggccgag taccccacga 1620
gagtcactcg atcacttcac tcgaccccca ggtcgagtgt cttcacctcc accacctgac 1680
catcactcga tcacgacgact ctacctggaa gtcgagtatc accatcacca ccactcgact 1740
acatacttga tgtcgagctt cagagtcttc tccattccgc actcaaccag acactcgagc 1800
acaaggaaaa aaagaagatt ctagcttatc actcgacctc tcactcgacc acctgggtcg 1860
agtacagttc ttaatccgtc tcaatactgc gtcgttttga gtattagggt ttcggaatat 1920
ttttgctata agtagcacgt actttacatt ttcgag 1956
```

```
<210> 8
<211> 2105
<212> DNA
<213> Arabidopsis thaliana

<400> 8
ttagcatttt ggttcactag ataaactcag tgccctcgtt gttagttgtc tgatgcatac 60
tcaatgaaat tgaagtaaaa ctgcaccatg cctagattgc tctactcgac cacactgtta 120
ggatctgata ccattcccta tcaatttgaa cttgaatctg atttagaatt atcatgtctt 180
gccatcgaat ttgaactcat ggatacccta aaatacttgg attttcttac tcattttaac 240
cactcttgtt aatccaagta gctgactctc cttattagag cagttaaccc gaatccaaac 300
ctaatctttc tttcgagccc tatatcactt gtgagtgttt gtgaggtctt atttcaattg 360
agcttggtag aaagtgttag gttcgtaacg acagagatag tgtctcatgt agttctagtt 420
cgcgtttttt ggactggata ggactaggtg ggcgcttata tcatgggttt ggatcagttt 480
aaaaaaaaaa aagggtgaat tcattgttga taaggaaagg gaaagaattc taggggaagt 540
aagctaaaga agttagaaaa aaaaaaatct agtaaaggtt ttgggaatgt taaagaaaag 600
aatgaggttc ttgttagcta aagaagaagg gttaaaagcc ttttgtttta aagattaaaa 660
acaggaacct tagttgttaa agaaatccaa atacgctaga tgtatcagag tgttgagaaa 720
gcttctccta gagttaagag aaaagaaaag aatgatatga aaaagagttt gaaagattca 780
tgagtgcaaa gggtagagtt aagttcttgt attgggactg gagttgggat taccattaga 840
gcttcattgt tatactatgg gtagatggga tttttatctct gtatgcataa cttgggactt 900
acctttagca ttctactaaa gctcaatcat tcttgagaga tcccctgtta cttaagccta 960
ttctgtaagg gaccatcttt gtctcttgcc ttcaccttag ccaaatgagt tcattgatga 1020
tgcattgttt gattcacgtt ccagaactaa tgaatgttaa agggattggt agatttgaaa 1080
gcatgtgtag gtcgagtata agagacggat tgattgataa caaggcatgg ctaacgtttt 1140
cgagtaaaat tcaatcatat cgcatcttag aactaccaac ttggacattg attttatttg 1200
ctctatcaga tgctttggtt ctgagtcccc accttcaaac ctctccttca actatgtctt 1260
cttatttgct tgagggcaag caaagactaa gtttggggga gttgatatgt ctataatttg 1320
catgttttca gtgtccattc atcatcgttt tgagtccagt ttcgtatcat tcatcactgt 1380
tttatatcat ttctcatcat tcttgcatac tttgcatgat taggatagct ttgtatacat 1440
attgcatttc tgagttgttt ttaggtgatt tggagctgtt tgcgagcaaa ttggaagaaa 1500
cgagccagaa caagaagcca tactcgaccc cctggtcgag tgactttgga gccattcttc 1560
ccatctactc gacccggggg tcgagtaacc tcagctcagg ccactcgatg acgccactcg 1620
tcccccctgg tcgagtatca cttcgccaca ccacctgacc acactcggcc gttcactcta 1680
ccacgttact cgaccccctg gtcgagtatc atcactcacc accaacacca tcactcgacc 1740
gggcactcga tcacatcttc atagtctact caaatccgca ctcaaccaga caagctgagc 1800
acaaggaaga gaagaggaga agacaaagtg tttggaagcg gcctggacct ccatcggatc 1860
acgaagaagc ccatctcggc ccattatcat tctatgggcc gggcgattag gttattggcc 1920
cgtctactat catttttattt cgttatgtat aaatagatgt cttagggttc tgtaccagga 1980
catctagtcg acattgagtt tttttgcttc agtttattt tctgttttct ctgctgcgcc 2040
gcttttgttt ctgcaacctg taattcgaga tttttccaag ttattcagat ccgcatttg 2100
atttc 2105
```

```
<210> 9
<211> 4860
<212> DNA
```

<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Virtual TSI

<400> 9
```
tcgccgttcc agcttcgcat actctctcac cgtctctcgt ttcactcgac cacttcacac 60
ttcgcctcaa catcttcgcc ggagtttctc gccattgtcc gtgcttccgt catctccgtt 120
cactcgacca ccggaccggc ttcaccatct ctcaactatc caccattcac tcgacctcgc 180
cattcactgc gcctccattc gtctctttac tcgactgctc ctcaaaccgc caccgtcttc 240
tctaaattcg ccgtttactc gaccacactg ttacgtctct cattcgtgta cagtcgaccg 300
ctatacccga agccacaata tcactctact cgaccgtttc actcgatcgc gtacttgact 360
ggtttagtgt gtgtgtttat ttgaactaac atattgatat ttggttttga gttacattct 420
ttttcaggga atcaatatga gcaactacag tggcgaatcc tccatggatg cggattacaa 480
cgtcgatgaa gctgaatctt ggtcaactag accagagaga gagcaacagg cttatgagag 540
cttcagagcc gagacccaac gctcagtagc tcgacgcaat gaaaggagag ctgagattgc 600
tagaggaaag agagcgatga ccagcagata tgagttgatc gacgaagata ttgacgtcga 660
gtatgagcct gagtcatggc acagagaaac aaaactgttg aacaagcctg atgaagttac 720
agtggaagag tacatcagac ttttcgagct gaacgacttc tggggagcga ggtacccctg 780
ttatgagact ctagcccagc ttaggctact ggaggacgta cagcacttat tcgagaagtg 840
ccatcttgag acgctgatgt cttacccgta cgtcgcttac aagaaggaaa caatagagtt 900
tctctccact ctgcaagtgg agttgtatca gggacttact gcagatgaac tggagagtga 960
agggttggga ttcttgactt tttcagttaa cgagcagcgt taccagctat ctatcaagag 1020
cttggaagga ttatttggtt ttcccagtgg aaagggaact aaacccaagt tcgaaaggga 1080
agagttgaag gatttgtggt taaccattgg gaacgatttg gcgctcaact ctgcaaggtc 1140
taagagcaac cagattcgaa gccctgtgat ccgctactat cagcgctcag tagcgaatgt 1200
tctgtacccc agggaatcta caggcaccgt gtctaacaca gacatggaga tgattgattc 1260
tgcactcaag ggtattctcc ggagaacaaa ggggaagaag gtcctaaagg gcgaccttaa 1320
tgatacacca ccggtcatgc ttctgttgat ccatatgtgt ggatacagga agtgggcgca 1380
caccaacggg aggaagaagg tgcgaggagc cctttgtgtg ggtggcgttg tgacaccgat 1440
tctgattgca tgtggtgtac ctctcacgtc tccagggttt gatccgagga tgatggattt 1500
agatcacttg cgtcgttgtg agtttctgga gtacgacatg gttggcgatt tctatcgcta 1560
caaattcgag cactccctga cccgaacagc caacattttg cttccctgca tcgaggccac 1620
aaccatactt tagggtgaga acattgactt cagacctgcg cgtgattacc tctactttga 1680
gagcactcca ccgactgatg acaatgtccc tacgacggaa gctacagagg atgattttgc 1740
tgagacggat gaggataggg aggaggagta tgatacgagc atgtatcatt tcagtgagca 1800
cgtacctcca gcgcaggaga gcaagagctt gagtgaagct cacagaaaca acagtaagtt 1860
gcagaggtgg tgcaagaaac aagataggtt acttatcaag tgcttcaagg ccatcacgtt 1920
tctaacggac aagataagtt gcttctcttc taccacagct attccgcagg gagagcgtcc 1980
tcaggacatg ccttcgaaga gatatgacgc gccagggcca agtcatcaca ggcctgagcc 2040
aagtcaccac aggcctgagc ctagtgaccg agtagtacca ccagtccctg caaggcattc 2100
atcattcgag cctcgggagc tcgggagaaa gaagaaggct gcactcgcta ggtctggcag 2160
caggagtaga cgacttctcc agtcccgtag cttacgcgac cgcggtgctg gccgcagcag 2220
aagaagagag gtcgagtatc atcagagcgg tgctggccgc ggcgaaggag cagaggtcga 2280
gtaccccag ggggaagctg agacacaaca gggagattct tcgatggcct gggagcaatc 2340
acaggcagct attgacgacc aactccgctc cttcttccac tgaggtatgc acctcactcc 2400
accattgtaa tataccatct cttgtttttt attttgtttt tgtgatgtgt tttgtcctga 2460
gtactctctt ccaaatttgg tcacacagtg gactgtgtga tttaagtttg ggggagggct 2520
caggaagtgt gtgttgcatt gtataatctt gagtttgcat tcatctaagg catagaaaaa 2580
ccaaaaaaat tgaaaaattc cagaaaatga tttcacaaaa atagagtgtt catgtagttg 2640
cattgcattt aggatcgagt ttagagtgtt tcgtttagga ttgttgcata tgcataggggg 2700
ataataatga gatagccttg taagcatttt ggttcaccag ataagctcag tgccctcgtt 2760
gttagttgtt tgatgcgttg tcattgaaat tgaagtaaga actgcacgat gcctagattg 2820
ctctactcga ccacactgtt aggatctgat atcattccct atcaatttga acttgaatct 2880
gatttagaat tatcatgtct tggcatcgaa tttgaactca tggataccct aaaatacttg 2940
gattttctta ctcatttaa ccactcttgt tgatccaagt agctgactct ccttattaga 3000
gcagttaacc catacccaaa cctgaacttt ctttcaagcc ctatatcact tgtgagtgtt 3060
tgtgaggtct tatttcgatt gagcttggta gaaagtgtta ggttcgtaac gacagagata 3120
gtgtctcatg tagttctagt ttgcgttctt cagactggat aggactaggt gggcgcttat 3180
atcatgggtt gggatgtgtt taaaagaaaa gagggaatct attgttgatg aggaaaggga 3240
aagaattcca ggggaagtaa gctaaagaag ttagaaaaaa aatctagtaa aggttttggg 3300
aatgttaaag aaaagaatga ggttcttgtt agctaaagaa gaagggttaa aagcctttgg 3360
ttttaaagat taaaaaaaaa acaggaacct tagttgttaa agaaatccaa acccgctaga 3420
```

31147

```
tgtatcaaga gcgttgagaa agcttctcct agagttaaga gaaaagaaaa gaatgatatg 3480
aaaaagagtt tgaaagattc atgagtgcaa agggtagagt taagttggga caggagttgg 3540
ttttaccatt agaacttcat tgttatactc tgggtagatg ggatcttatc tctgtatgca 3600
taatttggga cttaccttta gcattctact aaagctcaat cattcttgag ggatcccctg 3660
ttacttaagc ctattctata agggaccatc tttgtctctt gaccttcacc ttggccgaat 3720
gagttcattg atgatgcatt gcttgattcg cgttccagaa ctaatgaatg ttaaagggat 3780
tggtagattt gaaagcatgt gtaggtcgag tataagagac ggattgattg aaaacaaggc 3840
atggctaacg ttttttgagta gaattcaatc atatcgcatc ttagaactac caacttggac 3900
attgatttta tttgctctat catatgcttt ggttttgagt ccccgccttc actcctctcc 3960
ttcaactatg tcttcttatt tgcttgaggg caagcaaaga ctaagtttga gggagttgat 4020
atgtctataa tttgcatgtt ttcagtgtcc attcatcatc gtttttgagtc cagtttcgta 4080
tcattcatca ctgtttttata tcatttctca tcattcttgc atactttgca tgattaggat 4140
aactttgcat acatattgca tttctgagtt gttttcaggt gatttggagc tgtttgcaag 4200
caaattggaa gaaatgagcc agaaccagaa gacatactcg acccctaggt cgagtgactt 4260
tggggccatt cttcccacat actcggcccc caggtcgagt gactttggag ccattcttcc 4320
catccactcg accacgggt cgagtaacct tagctcaggc cactcgatga cactactcga 4380
ccccaggtcg agtatcactt cgccacacca cctgacaaca ctcgaccaat cactctacca 4440
agttactcga cccctggtc gagtatcatc actcaccacc atcagcatca ctcgaccgga 4500
cactcgatca cgtcttcaca gtctactcaa atccgcagtc aaccagacaa gctgagcaca 4560
aggaagagaa gaggagaaga caaagtgctt ggaagcggcc tggacctcca tcggatcacg 4620
aagcccatct cggcccatta tctctctatg ggccgagcga ttaggttatt ggcccgtcta 4680
ctatcatttt atttcgtttt gtataaatag atgtcttagg gttttgtcct gagacatcta 4740
gtcgacattg agtttttttt gcttcagttt tattttctgt tctactctgc tgcgccgctt 4800
ttgcttctgc aacctgtaat tcgagatttt tccaagttat tcagattccg catttgattt 4860
```

```
<210> 10
<211> 648
<212> PRT
<213> Arabidopsis thaliana

<400> 10
```

```
Met Ser Asn Tyr Ser Gly Glu Ser Ser Met Asp Ala Asp Tyr Asn Val
1               5                   10                  15

Asp Glu Ala Glu Ser Trp Ser Thr Arg Pro Glu Arg Glu Gln Gln Ala
            20                  25                  30

Tyr Glu Ser Phe Arg Ala Glu Thr Gln Arg Ser Val Ala Arg Arg Asn
        35                  40                  45

Glu Arg Arg Ala Glu Ile Ala Arg Gly Lys Arg Ala Met Thr Ser Arg
    50                  55                  60

Tyr Glu Leu Ile Asp Glu Asp Ile Asp Val Glu Tyr Glu Pro Glu Ser
65                  70                  75                  80

Trp His Arg Glu Thr Lys Leu Leu Asn Lys Pro Asp Glu Val Thr Val
                85                  90                  95

Glu Glu Tyr Ile Arg Leu Phe Glu Leu Asn Asp Phe Trp Gly Ala Arg
            100                 105                 110

Tyr Pro Cys Tyr Glu Thr Leu Ala Gln Leu Arg Leu Leu Glu Asp Val
        115                 120                 125

Gln His Leu Phe Glu Lys Cys His Leu Glu Thr Leu Met Ser Tyr Pro
    130                 135                 140

Tyr Val Ala Tyr Lys Lys Glu Thr Ile Glu Phe Leu Ser Thr Leu Gln
145                 150                 155                 160

Val Glu Leu Tyr Gln Gly Leu Thr Ala Asp Glu Leu Glu Ser Glu Gly
```

Page 7

```
                        165                      170              175
Leu Gly Phe Leu Thr Phe Ser Val Asn Glu Gln Arg Tyr Gln Leu Ser
            180                 185                 190

Ile Lys Ser Leu Glu Gly Leu Phe Gly Phe Pro Ser Gly Lys Gly Thr
            195                 200                 205

Lys Pro Lys Phe Glu Arg Glu Glu Leu Lys Asp Leu Trp Leu Thr Ile
    210                 215                 220

Gly Asn Asp Leu Ala Leu Asn Ser Ala Arg Ser Lys Ser Asn Gln Ile
225                 230                 235                 240

Arg Ser Pro Val Ile Arg Tyr Tyr Gln Arg Ser Val Ala Asn Val Leu
                245                 250                 255

Tyr Pro Arg Glu Ser Thr Gly Thr Val Ser Asn Thr Asp Met Glu Met
            260                 265                 270

Ile Asp Ser Ala Leu Lys Gly Ile Leu Arg Arg Thr Lys Gly Lys Lys
            275                 280                 285

Val Leu Lys Gly Asp Leu Asn Asp Thr Pro Pro Val Met Leu Leu Leu
    290                 295                 300

Ile His Met Cys Gly Tyr Arg Lys Trp Ala His Thr Asn Gly Arg Lys
305                 310                 315                 320

Lys Val Arg Gly Ala Leu Cys Val Gly Gly Val Val Thr Pro Ile Leu
                325                 330                 335

Ile Ala Cys Gly Val Pro Leu Thr Ser Pro Gly Phe Asp Pro Arg Met
            340                 345                 350

Met Asp Leu Asp His Leu Arg Arg Cys Glu Phe Leu Glu Tyr Asp Met
            355                 360                 365

Val Gly Asp Phe Tyr Arg Tyr Lys Phe Glu His Ser Leu Thr Arg Thr
    370                 375                 380

Ala Asn Ile Leu Leu Pro Cys Ile Glu Ala Thr Thr Ile Leu Xaa Gly
385                 390                 395                 400

Glu Asn Ile Asp Phe Arg Pro Ala Arg Asp Tyr Leu Tyr Phe Glu Ser
                405                 410                 415

Thr Pro Pro Thr Asp Asp Asn Val Pro Thr Thr Glu Ala Thr Glu Asp
            420                 425                 430

Asp Phe Ala Glu Thr Asp Glu Asp Arg Glu Glu Glu Tyr Asp Thr Ser
    435                 440                 445

Met Tyr His Phe Ser Glu His Val Pro Pro Ala Gln Glu Ser Lys Ser
    450                 455                 460

Leu Ser Glu Ala His Arg Asn Asn Ser Lys Leu Gln Arg Trp Cys Lys
465                 470                 475                 480

Lys Gln Asp Arg Leu Leu Ile Lys Cys Phe Lys Ala Ile Thr Phe Leu
                485                 490                 495

Thr Asp Lys Ile Ser Cys Phe Ser Ser Thr Thr Ala Ile Pro Gln Gly
                              Page 8
```

31147

```
                             500                     505                     510
Glu Arg Pro Gln Asp Met Pro Ser Lys Arg Tyr Asp Ala Pro Gly Pro
        515                 520                 525

Ser His His Arg Pro Glu Pro Ser His His Arg Pro Glu Pro Ser Asp
    530                 535                 540

Arg Val Val Pro Pro Val Pro Ala Arg His Ser Ser Phe Glu Pro Arg
545                 550                 555                 560

Glu Leu Gly Arg Lys Lys Lys Ala Ala Leu Ala Arg Ser Gly Ser Arg
                565                 570                 575

Ser Arg Arg Leu Leu Gln Ser Arg Ser Leu Arg Asp Arg Gly Ala Gly
            580                 585                 590

Arg Ser Arg Arg Arg Glu Val Glu Tyr His Gln Ser Gly Ala Gly Arg
        595                 600                 605

Gly Glu Gly Ala Glu Val Glu Tyr Pro Gln Gly Glu Ala Glu Thr Gln
    610                 615                 620

Gln Gly Asp Ser Ser Met Ala Trp Glu Gln Ser Gln Ala Ala Ile Asp
625                 630                 635                 640

Asp Gln Leu Arg Ser Phe Phe His
                645
```

```
<210> 11
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 11
tggttcacca gataagctca gtgccctc                                        28


<210> 12
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 12
cttcagactg gataggacta ggtgggcg                                        28


<210> 13
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
```

31147

```
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 13
cgcccaccta gtcctatcca gtctgaag                                28


<210> 14
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 14
cgcatcaaac aactaacaac gagggcac                                28


<210> 15
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 15
cgataacatc gaccgtattg ctcgcc                                  26


<210> 16
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 16
aactagctcc catccgtctt cgacatcc                                28


<210> 17
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
      Oligonucleotide sequence

<400> 17
tgcatcacac cggattggat tgac                                    24


<210> 18
<211> 28
<212> DNA
<213> Artificial Sequence
```

31147

<220>
<223> Description of Artificial Sequence:
     Oligonucleotide sequence

<400> 18
tgttcccctg aaccatagca atgagacc   28

<210> 19
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     Oligonucleotide sequence

<400> 19
caaacagaca gagtgtggcc caccacc   27

<210> 20
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     Oligonucleotide sequence

<400> 20
caaacagaca gagtgtggcc caccacc   27

<210> 21
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     Oligonucleotide sequence

<400> 21
tgcaaaccca caggaccaag tctaccc   27

<210> 22
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     Oligonucleotide sequence

<400> 22
acagatggtg atagcgtgag cggtggc   27

<210> 23
<211> 26

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     oligonucleotide sequence

<400> 23
tcaacctttt gccccaacaa ccactc   26

<210> 24
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     oligonucleotide sequence

<400> 24
tctccatcca cgctttcctg aatgtcc   27

<210> 25
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     oligonucleotide sequence

<400> 25
ggagaaggaa gctgaaaatc atattgtgg   29

<210> 26
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence:
     oligonucleotide sequence

<400> 26
atgatgatcc taagtctacc cttttgcac   29

<210> 27
<211> 3122
<212> DNA
<213> Arabidopsis thaliana

<400> 27
ttcatatatt cgacctcttc ttctcattct tgcatccaaa agacacaaca agccgccatc 60
gctttccctc acaactctca ctcgaccacc gcccgctctc tcacttactc ggcttcatcg 120
ctctcatcgc catctctcaa catactcgac ctcgcgatat cactcgagct cgccgcttct 180
caccgcctct ccatcgtcac cgcctgctcc ctctctccaa ggaaacaact cgagctctcc 240
atttcactca ctcgacctct accaccaagc cggcttcacc acttctagct cttaaccact 300
cgaccacctt caccatcaac caatcaaatc gttttctcct ccattaaagc ttgacatact 360
cgaccgctga acacttatca ccttcaagct cctcatctct tcatcgtttc caacaccgct 420

31147

```
gctctcatcc cccacgaaag cttgtcatca cctctcactc atcaccagtt cactcgattc 480
agcaaccaaa ctcgacctcg tctctcttgc cactcatagt cactcgatct ctcctcacca 540
tcttcatcat ctcccttact cgaccaccgt gcgtctcgct ccaccattgc catttaaaag 600
ctcactcgat tgtcaaagag aagaagagtg aagctcaacc accgccactc gaccgcgttt 660
ccctctacac attcaacact cgaccacggt gctaccatct ccacacccgc tcttgttcac 720
catacactcg accaacaact ctcaaagtaa aaaaaaaaag aaaaaaaaag tcaaaaccga 780
cagtttcact caaccggttt actcgaccgg tacgctggtt tagattgtgt ttttggtttt 840
gctattacta acatattaac gtttatcttt gagtttcgtc tgttttagg tttcatcatg 900
agtaactaca gtggaaaatc ctctatggac cctgattata atgtggatga agctaagtcc 960
tggtccacta gaccggagtg agagcaacat gtttacgaga gctataggga tgaatttgaa 1020
cgctctgcag ctcgacgtaa tcaaagaaga gctgaaatcg ctagaggaaa gagggcgatg 1080
tcgagtagat atgagctgat tgatgaggat atcaaaactg agtatgagcc agagtcatgg 1140
cgcaaggaga cgaagctact gaacaaatcc gacgaggtta cagtggagga gtatatcaga 1200
ttctttgaga tgaatgactt ctggggaacg aggtatccct gatatgagac tttagcccag 1260
ttgggggttac tggaggacgt gcagcatctg ttcgagaagt gtcatctgat aaggaggaga 1320
caatcgagtt tctttccaca ctgcaagtgg aaatgtatga gggactcaca gactttgagc 1380
tggataccat ggggttaggc ttcttgacgt tcttagtgga tgaacagcgg taccagattt 1440
agatcaagaa attggaagaa ctgtttggtt tccctagtgg aaagggaacc aaccccaggt 1500
ttgacaggga gagcttaag gatttgtggg ctactattgg gaacaatcta ccgctaaact 1560
cgacgcggtc caagagcaac caaatccgga gtcctgtgat tcgctacttt cagcgctcgg 1620
ttgccaatgt tttttactcc agggagtcta caggcaccgt gtctaacaca gacatgaaga 1680
tgatagattc agcgcttata gggattctcc gccttacaaa aggaaagaat gtcctgagag 1740
gagatcttaa cgactcacca ccagtaatgc ctctgttgat ccatctgtgt gggtacatga 1800
agtgggcgct gacaaacggc aagaagaagg taagaggagc actatgcgtg ggtggcgttg 1860
tgacgccaat tctgaaagtt tgtggagttc cgctcaagga agtagggtta gcaccagaa 1920
tgatggactt ggatcacttg cgccgatgtg agttctctga gtttgacatg gttggcgact 1980
ttcaccgcta caggttcgag cattcatcga ttagaatcgc caacattctt ttcccctgca 2040
tttacgctac taggattctc gagggcagga acattgactt caagcctgcg cttgaagatc 2100
tttatttcga gggcagtccg ccaactgagg agattagtca caccgaagga gctacaatag 2160
aagatgttga tgagacatat gatatagatg aggcggagtt tgacacgagc atgtatcatt 2220
tcagtgagca tatacctcca gcgaggaaaa gcaagagttt gagcgaagct cacaggaaca 2280
acagcaagct gcagaagtgg tgcaagaaac aggataagtt actcgccaag tgcctcaggg 2340
ctatcaagtt tctgaaggac aagatcagct gctcctcttc cactacaact attccgcaat 2400
gacagctccc tcaggacatg ccttcgagga gatatgacgc gcccgagcct agagagcaga 2460
agattctgca tgtccctgcg aggcattcat cattcgaatct tcgtgaatct aggaagaata 2520
ggagaacgac actcactcga tctagcagca ggagcagacg acttctgcag tctcgtagtt 2580
tacgcgaccg cggtgctggc cgcaatagaa gaagagaggt cgagtatcct cagagcggtg 2640
ctggccgcca cagagctgat gagatcgagt acccacatgc tggagctgat acggaacatg 2700
gcggttcgtc tatggcttgg gagcaatcac aggcagccat tgactaccaa cttcgttcat 2760
tattcgactg aggtaagcgc ctcacttcac cattatatta tatcatctct tgtgatttgt 2820
tctttatttt gtttcagtga ttggatttgt cctgagtact ctcttccaag tttattcaca 2880
cagtggactg tgtgatttaa gtttgggggga gggctcagga agtatgttgc attgtatata 2940
tttttaagtc tgcattcatc taaggcatag aaaaaccaaa aaaaaattaa aaatttcaga 3000
aaatgatttc acaaaaaaag agtgttcatg tagttgcatt acatttagga tcaagtctag 3060
agtgtttcat ttaggattgt tgcatatgca taggggataa tgatgagata gccttgtaag 3120
ca                                                                3122
```